FIGURE 1

Plasmid sequence of pNC5LSPCEAp53 (pMC30B5) for vCP2086

```
        1  GCCCTTT CGTCTCG CGCGTTT CGGTGAT GACGGTG AAAACCT CTGACAC ATGCAGC TCCCGGA GACGGTC
           CGGGAAA GCAGAGC GCGCAAA GCCACTA CTGCCAC TTTTGGA GACTGTG TACGTCG AGGGCCT CTGCCAG
       71  ACAGCTT GTCTGTA AGCGGAT GCCGGGA GCAGACA AGCCCGT CAGGGCG CGTCAGC GGGTGTT GGCGGGT
           TGTCGAA CAGACAT TCGCCTA CGGCCCT CGTCTGT TCGGGCA GTCCCGC GCAGTCG CCCACAA CCGCCCA
      141  GTCGGGG CTGGCTT AACTATG CGGCATC AGAGCAG ATTGTAC TGAGAGT GCACCAT ATGCGGT GTGAAAT
           CAGCCCC GACCGAA TTGATAC GCCGTAG TCTCGTC TAACATG ACTCTCA CGTGGTA TACGCCA CACTTTA
      211  ACCGCAC AGATGCG TAAGGAG AAAATAC CGCATCA GGCGCCA TTCGCCA TTCAGGC TGCGCAA CTGTTGG
           TGGCGTG TCTACGC ATTCCTC TTTTATG GCGTAGT CCGCGGT AAGCGGT AAGTCCG ACGCGTT GACAACC
      281  GAAGGGC GATCGGT GCGGGCC TCTTCGC TATTACG CCAGCTG GCGAAAG GGGGATG TGCTGCA AGGCGAT
           CTTCCCG CTAGCCA CGCCCGG AGAAGCG ATAATGC GGTCGAC CGCTTTC CCCCTAC ACGACGT TCCGCTA
      351  TAAGTTG GGTAACG CCAGGGT TTTCCCA GTCACGA CGTTGTA AAACGAC GGCCAGT GCCAAGC TTGGCTG
           ATTCAAC CCATTGC GGTCCCA AAAGGGT CAGTGCT GCAACAT TTTGCTG CCGGTCA CGGTTCG AACCGAC
           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                                  Left Arm
      421  CAGGTAT TCTAAAC TAGGAAT AGATGAA ATTATGT GCAAAGG AGATACC TTTAGAT ATGGATC TGATTTA
           GTCCATA AGATTTG ATCCTTA TCTACTT TAATACA CGTTTCC TCTATGG AAATCTA TACCTAG ACTAAAT
                                          Left Arm
      491  TTTGGTT TTTCATA ATCATAA TCTAACA ACATTTT CACTATA CTATACC TTCTTGC ACAAGTC GCCATTA
           AAACCAA AAAGTAT TAGTATT AGATTGT TGTAAAA GTGATAT GATATGG AAGAACG TGTTCAG CGGTAAT
                                          Left Arm
      561  GTAGTAT AGACTTA TACTTTG TAACCAT AGTATAC TTTAGCG CGTCATC TTCTTCA TCTAAAA CAGATTT
           CATCATA TCTGAAT ATGAAAC ATTGGTA TCATATG AAATCGC GCAGTAG AAGAAGT AGATTTT GTCTAAA
                                          Left Arm
      631  ACAACAA TAATCAT CGTCGTC ATCTTCA TCTTCAT TAAAGTT TTCATAT TCAATAA CTTTCTT TTCTAAA
           TGTTGTT ATTAGTA GCAGCAG TAGAAGT AGAAGTA ATTTCAA AAGTATA AGTTATT GAAAGAA AAGATTT
                                          Left Arm
      701  ACATCAT CTGAATC AATAAAC ATAGAAC GGTATAG AGCGTTA ATCTCCA TTGTAAA ATATACT AACGCGT
           TGTAGTA GACTTAG TTATTTG TATCTTG CCATATC TCGCAAT TAGAGGT AACATTT TATATGA TTGCGCA
                                          Left Arm
      771  TGCTCAT GATGTAC TTTTTTT CATTATT TAGAAAT TATGCAT TTTAGAT CTTTATA AGCGGCC GTGATTA
           ACGAGTA CTACATG AAAAAAA GTAATAA ATCTTTA ATACGTA AAATCTA GAAATAT TCGCCGG CACTAAT
                                                                          ~~~
                                      Left Arm
      841  ACTAGTC ATAAAAA CCCGGGA TCGATTC TAGACTC GAGATAA AAACTAT ATCAGAG CAACCCC AACCAGC
           TGATCAG TATTTTT GGGCCCT AGCTAAG ATCTGAG CTCTATT TTTGATA TAGTCTC GTTGGGG TTGGTCG
           ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                  CEA
                                                    ***Ile LeuAla ValGly ValLeuVal·
      911  ACTCCAA TCATGAT GCCGACA GTGGCCC CAGCTGA GAGACCA GGAGAAG TTCCAGA TGCAGAG ACTGTGA
           TGAGGTT AGTACTA CGGCTGT CACCGGG GTCGACT CTCTGGT CCTCTTC AAGGTCT ACGTCTC TGACACT
                                      CEA
           ..GlyIle MetIle GlyValThr AlaGly AlaSer LeuGlyPro SerThr GlySer AlaSerVal ThrIle·
      981  TGCTCTT GACTATG GAATTAT TGCGGCC AGTAGCC AAGTTAG AGACAAA ACAGGCA TAGGTCC CGTTATT
           ACGAGAA CTGATAC CTTAATA ACGCCGG TCATCGG TTCAATC TCTGTTT TGTCCGT ATCCAGG GCAATAA
                                      CEA
           .SerLys ValIleSer AsnAsn ArgGly ThrAlaLeu AsnSer ValPhe CysAlaTyr ThrGly AsnAsn
     1051  ATTTGGC GTGATTT TGGCGAT AAAGAGA ACTTGTG TGTGTTG CTGCGGT ATCCCAT TGATACG CCAAGAA
           TAAACCG CACTAAA ACCGCTA TTTCTCT TGAACAC ACACAAC GACGCCA TAGGGTA ACTATGC GGTTCTT
                                      CEA
           AsnProThr IleLys AlaIle PheLeuVal GlnThr HisGln GlnProIle GlyAsn IleArg TrpSerTyr·
     1121  TACTGCG GGGATGG GTTAGAG GCCGAGT GGCAGGA GAGGTTG AGGTCCG CTCCCGA AAGGTAA GACGAGT
           ATGACGC CCCTACC CAATCTC CGGCTCA CCGTCCT CTCCAAC TCCAGGC GAGGGCT TTCCATT CTGCTCA
                                      CEA
           ..GlnPro SerPro AsnSerAla SerHis CysSer LeuAsnLeu AspAla GlySer LeuTyrSer SerAsp·
     1191  CTGGGGG GGAAATG ATGGGGG TGTCCGG CCCATAG AGGACAT CCAGGGT GACTGGG TCACTGC GGTTTGC
           GACCCCC CCTTTAC TACCCCC ACAGGCC GGGTATC TCCTGTA GGTCCCA CTGACCC AGTGACG CCAAACG
                                      CEA
           .ProPro SerIleIle ProThr AspPro GlyTyrLeu ValAsp LeuThr ValProAsp SerArg AsnAla
     1261  ACTCACT GAGTTCT GGATTCC ACATACA TAGGCTC TTGCGTC ATTTCTT GTGACAT TGAATAG AGTGAGG
           TGAGTGA CTCAAGA CCTAAGG TGTATGT ATCCGAG AACGCAG TAAAGAA CACTGTA ACTTATC TCACTCC
                                      CEA
           SerValSer AsnGln IleGly CysValTyr AlaArg AlaAsp AsnArgThr ValAsn PheLeu ThrLeuThr·
     1331  GTCCTGT TGCCATT GGACAGC TGCAGCC TGGGACT GACTGGG AGGCTCT GACCATT TACCCAC CACAGGT
           CAGGACA ACGGTAA CCTGTCG ACGTCGG ACCCTGA CTGACCC TCCGAGA CTGGTAA ATGGGTG GTGTCCA
                                      CEA
           ..ArgAsn GlyAsn SerLeuGln LeuArg ProSer ValProLeu SerGln GlyAsn ValTrpTrp LeuTyr·
     1401  AGGTTGT GTTCTGA GCCTCAG GTTCACA GGTGAAG GCCACAG CATCCTT GTCCTCC ACGGGTT TGGAGTT
           TCCAACA CAAGACT CGGAGTC CAAGTGT CCACTTC CGGTGTC GTAGGAA CAGGAGG TGCCCAA ACCTCAA
                                      CEA
           .ThrThr AsnGlnAla GluPro GluCys ThrPheAla ValAla AspLys AspGluVal ProLys SerAsn
```

```
      1471  GTTGCTG GAGATGG AGGGCTT GGGCAGC TCCGCGG AAACAGT TATTGTT TTAACTG TAGTCCT GCTGTGA
            CAACGAC CTCTACC TCCCGAA CCCGTCG AGGCGCC TTTGTCA ATAACAA AATTGAC ATCAGGA CGACACT
                                                CEA
            AsnSerSer IleSer ProLys ProLeuGlu AlaSer ValThr IleThrLys ValThr ThrArg SerHisGly·
      1541  CCACTGG CTGAGTT ATTGGCC TGGCAAG TATAGAG TCCGCTG TTCTTCT CAGTTAT GTTGCTT ATAAATA
            GGTGACC GACTCAA TAACCGG ACCGTTC ATATCTC AGGCGAC AAGAAGA GTCAATA CAACGAA TATTTAT
                                                CEA
            ..SerAla SerAsn AsnAlaGln CysThr TyrLeu GlySerAsn LysGlu ThrIle AsnSerIle PheLeu·
      1611  ACTCTTG AGTATGC TGCTGAA TGTTTCC ATCAATC AGCCAGG AGTACTG TGCAGGG GGGTTGG ATGCTGC
            TGAGAAC TCATACG ACGACTT ACAAAGG TAGTTAG TCGGTCC TCATGAC ACGTCCC CCCAACC TACGACG
                                                CEA
            .GluGln ThrHisGln GlnIle AsnGly AspIleLeu TrpSer TyrGln AlaProPro AsnSer AlaAla
      1681  ATGGCAA GAAAGGC TCAAGTT CACGCCG GGACGGT AGTAGGT GTATGAT GGAGATA TAGTTGG GTCGTCT
            TACCGTT CTTTCCG AGTTCAA GTGCGGC CCTGCCA TCATCCA CATACTA CCTCTAT ATCAACC CAGCAGA
                                                CEA
            HisCysSer LeuSer LeuAsn ValGlyPro ArgTyr TyrThr TyrSerPro SerIle ThrPro AspAspPro·
      1751  GGGCCAT ACAAAAC ATTAAGG ATAACAG GGTCGGA GTGATCA ACGGATA ATTCATT CTGAATG CCACACT
            CCCGGTA TGTTTTG TAATTCC TATTGTC CCAGCCT CACTAGT TGCCTAT TAAGTAA GACTTAC GGTGTGA
                                                CEA
            ..GlyTyr LeuVal AsnLeuIle ValPro AspSer HisAspVal SerLeu GluAsn GlnIleGly CysGlu·
      1821  CATAAGG TCCTACA TCATTGC GAGTAAC GGACAGG AGTGTCA ATGTGCG GTTATCA TTAGACA ACTGCAA
            GTATTCC AGGATGT AGTAACG CTCATTG CCTGTCC TCACAGT TACACGC CAATAGT AATCTGT TGACGTT
                                                CEA
            .TyrPro GlyValAsp AsnArg ThrVal SerLeuLeu ThrLeu ThrArg AsnAspAsn SerLeu GlnLeu
      1891  GCGTGGG CTAACCG GCAAACT TTGGTTA TTGACCC ACCATAA ATAAGTG GTATTTT GAATCTC TGGCTCA
            CGCACCC GATTGGC CGTTTGA AACCAAT AACTGGG TGGTATT TATTCAC CATAAAA CTTAGAG ACCGAGT
                                                CEA
            ArgProSer ValPro LeuSer GlnAsnAsn ValTrp TrpLeu TyrThrThr AsnGln IleGlu ProGluCys·
      1961  CAAGTTA ATGCAAC TGCGTCC TCATCCT CAACTGG GTTAGAA TTGTTAC TAGTTAT GAATGGT TTTGGTG
            GTTCAAT TACGTTG ACGCAGG AGTAGGA GTTGACC CAATCTT AACAATG ATCAATA CTTACCA AAACCAC
                                                CEA
            ..ThrLeu AlaVal AlaAspGlu AspGlu ValPro AsnSerAsn AsnSer ThrIle PheProLys ProPro·
      2031  GCTCATA CACGGTA ATCGTCG TCACGGT TGTGCGG TTGAGTC CGGTGTC GCTATTG TGAGCTT GGCACGT
            CGAGTAT GTGCCAT TAGCAGC AGTGCCA ACACGCC AACTCAG GCCACAG CGATAAC ACTCGAA CCGTGCA
                                                CEA
            .GluTyr ValThrIle ThrThr ValThr ThrArgAsn LeuGly ThrAsp SerAsnHis AlaGln CysThr
      2101  GTAGGAT CCACTAT TGTTCAC GGTAATA TTGGGAA TGAACAG TTCCTGG GTGGACT GTTGGAA AGTGCCA
            CATCCTA GGTGATA ACAAGTG CCATTAT AACCCTT ACTTGTC AAGGACC CACCTGA CAACCTT TCACGGT
                                                CEA
            TyrSerGly SerAsn AsnVal ThrIleAsn ProIle PheLeu GluGlnThr SerGln GlnPhe ThrGlyAsn·
      2171  TTGACAA ACCAGCT GTATTGG GCGGGAG GATTGCT AGCGGCA TGACAGC TCAGATT CAGATTT TCCCCTG
            AACTGTT TGGTCGA CATAACC CGCCCTC CTAACGA TCGCCGT ACTGTCG AGTCTAA GTCTAAA AGGGGAC
                                                CEA
            ..ValPhe TrpSer TyrGlnAla ProPro AsnSer AlaAlaHis CysSer LeuAsn LeuAsnGlu GlySer·
      2241  ATCTATA GCTTGTG TTTAGAG GGCTGAT TGTAGGA GCATCGG GTCCGTA AAGCACG TTGAGAA TCACTGA
            TAGATAT CGAACAC AAATCTC CCGACTA ACATCCT CGTAGCC CAGGCAT TTCGTGC AACTCTT AGTGACT
                                                CEA
            .ArgTyr SerThrAsn LeuPro SerIle ThrProAla AspPro GlyTyr LeuValAsn LeuIle ValSer
      2311  ATCAGAC CTCCTGG CGCTGAC TGGATTT TGGGTTT CGCATTT GTAGCTT GCTGTGT CGTTCCT GGTCACG
            TAGTCTG GAGGACC GCGACTG ACCTAAA ACCCAAA GCGTAAA CATCGAA CGACACA GCAAGGA CCAGTGC
                                                CEA
            AspSerArg ArgAla SerVal ProAsnGln ThrGlu CysLys TyrSerAla ThrAsp AsnArg ThrValAsn·
      2381  TTAAACA GGGTCAG AGTTCTA TTTCCGT TGCTGAG TTGGAGT CTAGGGG ACACAGG CAGGGAC TGGTTGT
            AATTTGT CCCAGTC TCAAGAT AAAGGCA ACGACTC AACCTCA GATCCCC TGTGTCC GTCCCTG ACCAACA
                                                CEA
            ..PheLeu ThrLeu ThrArgAsn GlyAsn SerLeu GlnLeuArg ProSer ValPro LeuSerGln AsnAsn·
      2451  TCACCCA CCAGAGA TATGTTG CGTCTTG AGTTTCG GGCTCGC ATGTAAA AGCGACG GCATCTT TGTCTTC
            AGTGGGT GGTCTCT ATACAAC GCAGAAC TCAAAGC CCGAGCG TACATTT TCGCTGC CGTAGAA ACAGAAG
                                                CEA
            .ValTrp TrpLeuTyr ThrAla AspGln ThrGluPro GluCys ThrPhe AlaValAla AspLys AspGlu
      2521  GACAGGC TTACTAT TATTGGA GCTAATA GAAGGCT TAGGGAG TTCCGGG TATACCC GGAACTG GCCAGTT
            CTGTCCG AATGATA ATAACCT CGATTAT CTTCCGA ATCCCTC AAGGCCC ATATGGG CCTTGAC CGGTCAA
                                                CEA
            ValProLys SerAsn AsnSer SerIleSer ProLys ProLeu GluProTyr ValArg PheGln GlyThrAla·
      2591  GCTTCTT CATTCAC AAGATCT GACTTTA TGACGTG TAGGGTG TAGAATC CTGTGTC ATTCTGG ATGATGT
            CGAAGAA GTAAGTG TTCTAGA CTGAAAT ACTGCAC ATCCCAC ATCTTAG GACACAG TAAGACC TACTACA
                                                CEA
            ..GluGlu AsnVal LeuAspSer LysIle ValHis LeuThrTyr PheGly ThrAsp AsnGlnIle IleAsn·
      2661  TCTGGAT CAGCAGG GATGCAT TGGGGTA TATTATC TCTCGAC CACTGTA TGCGGGC CCTGGGG TAGCTTG
            AGACCTA GTCGTCC CTACGTA ACCCCAT ATAATAG AGAGCTG GTGACAT ACGCCCG GGACCCC ATCGAAC
                                                CEA
            .GlnIle LeuLeuSer AlaAsn ProTyr IleIleGlu ArgGly SerTyr AlaProGly ProThr AlaGln
      2731  TTGAGTT CCTATTA CATATCC TATAATT TGACGGT TGCCATC CACTCTT TCACCTT TGTACCA GCTGTAG
            AACTCAA GGATAAT GTATAGG ATATTAA ACTGCCA ACGGTAG GTGAGAA AGTGGAA ACATGGT CGACATC
```

```
                                        CEA
          GlnThrGly IleVal TyrGly IleIleGln ArgAsn GlyAsp ValArgGlu GlyLys TyrTrp SerTyrGly·
   2801   CCAAAAA GATGCTG GGGCAGA TTGTGGA CAAGTAG AAGCACC TCCTTCC CCTCTGC GACATTG AACGGCG
          GGTTTTT CTACGAC CCCGTCT AACACCT GTTCATC TTCGTGG AGGAAGG GGAGACG CTGTAAC TTGCCGC
                                        CEA
          ..PheLeu HisGln ProLeuAsn HisVal LeuLeu LeuValGlu LysGly GluAla ValAsnPhe ProThr·
   2871   TGGATTC AATAGTG AGCTTGG CAGTGGT GGGCGGG TTCCAGA AGGTTAG AAGTGAG GCTGTGA GCAGGAG
          ACCTAAG TTATCAC TCGAACC GTCACCA CCCGCCC AAGGTCT TCCAATC TTCACTC CGACACT CGTCCTC
                                        CEA
          .SerGlu IleThrLeu LysAla ThrThr ProProAsn TrpPhe ThrLeu LeuSerAla ThrLeu LeuLeu
   2941   CCTCTGC CAGGGGA TGCACCA TCTGTGG GGAGGGG CCGAGGG AGACTCC ATTATTT ATATTCC AAAAAAA
          GGAGACG GTCCCCT ACGTGGT AGACACC CCTCCCC GGCTCCC TCTGAGG TAATAAA TATAAGG TTTTTTT
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                         E/L Promoter
                                                                ~~~~~~~~~~~~~~~~~~~~
                               CEA
          ArgGlnTrp ProIle CysTrp ArgHisPro ProAla SerPro SerGluMet
                                                                          H6 promoter
                                                                          ~~~~~~~~~~
   3011   AAAAATA AAATTTC AATTTTT GTCGACC TGCAGCT CGACGGA TCCCCCC GGGTTCT TTATTCT ATACTTA
          TTTTTAT TTTAAAG TTAAAAA CAGCTGG ACGTCGA GCTGCCT AGGGGGG CCCAAGA AATAAGA TATGAAT
                               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
             E/L Promoter
                                  H6 promoter
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   3081   AAAAGTG AAAATAA ATACAAA GGTTCTT GAGGGTT GTGTTAA ATTGAAA GCGAGAA ATAATCA TAAATTA
          TTTTCAC TTTTATT TATGTTT CCAAGAA CTCCCAA CACAATT TAACTTT CGCTCTT TATTAGT ATTTAAT
                                                                  p53
                                                    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                   H6 promoter
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                    MetGlu GluProGln SerAsp ProSer ValGluPro·
   3151   TTTCATT ATCGCGA TATCCGT TAAGTTT GTATCGT AATGGAG GAGCCGC AGTCAGA TCCTAGC GTCGAGC
          AAAGTAA TAGCGCT ATAGGCA ATTCAAA CATAGCA TTACCTC CTCGGCG TCAGTCT AGGATCG CAGCTCG
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          ..ProLeu SerGln GluThrPhe SerAsp LeuTrp LysLeuLeu ProGlu AsnAsn ValLeuSer ProLeu·
   3221   CCCCTCT GAGTCAG GAAACAT TTTCAGA CCTATGG AAACTAC TTCCTGA AAACAAC GTTCTGT CCCCCTT
          GGGGAGA CTCAGTC CTTTGTA AAAGTCT GGATACC TTTGATG AAGGACT TTTGTTG CAAGACA GGGGGAA
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          .ProSer GlnAlaMet AspAsp LeuMet LeuSerPro AspAsp IleGlu GlnTrpPhe ThrGlu AspPro
   3291   GCCGTCC CAAGCAA TGGATGA TTTGATG CTGTCCC CGGACGA TATTGAA CAATGGT TCACTGA AGACCCA
          CGGCAGG GTTCGTT ACCTACT AAACTAC GACAGGG GCCTGCT ATAACTT GTTACCA AGTGACT TCTGGGT
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          GlyProAsp GluAla ProArg MetProGlu AlaAla ProPro ValAlaPro AlaPro AlaAla ProThrPro·
   3361   GGTCCAG ATGAAGC TCCCAGA ATGCCAG AGGCTGC TCCCCCC GTGGCCC CTGCACC AGCAGCT CCTACAC
          CCAGGTC TACTTCG AGGGTCT TACGGTC TCCGACG AGGGGGG CACCGGG GACGTGG TCGTCGA GGATGTG
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          ..AlaAla ProAla ProAlaPro SerTrp ProLeu SerSerSer ValPro SerGln LysThrTyr GlnGly·
   3431   CGGCGGC CCCTGCA CCAGCCC CCTCCTG GCCCCTG TCATCTT CTGTCCC TTCCCAG AAAACCT ACCAGGG
          GCCGCCG GGGACGT GGTCGGG GGAGGAC CGGGGAC AGTAGAA GACAGGG AAGGGTC TTTTGGA TGGTCCC
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          .SerTyr GlyPheArg LeuGly PheLeu HisSerGly ThrAla LysSer ValThrCys ThrTyr SerPro
   3501   CAGCTAC GGTTTCC GTCTGGG CTTCTTG CATTCTG GGACAGC CAAGTCT GTGACTT GCACGTA CTCCCCT
          GTCGATG CCAAAGG CAGACCC GAAGAAC GTAAGAC CCTGTCG GTTCAGA CACTGAA CGTGCAT GAGGGGA
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          AlaLeuAsn LysMet PheCys GlnLeuAla LysThr CysPro ValGlnLeu TrpVal AspSer ThrProPro·
   3571   GCCCTCA ACAAGAT GTTTTGC CAACTGG CCAAGAC CTGCCCT GTGCAGC TGTGGGT TGATTCC ACACCCC
          CGGGAGT TGTTCTA CAAAACG GTTGACC GGTTCTG GACGGGA CACGTCG ACACCCA ACTAAGG TGTGGGG
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          ..ProGly ThrArg ValArgAla MetAla IleTyr LysGlnSer GlnHis MetThr GluValVal ArgArg·
   3641   CGCCCGG CACCCGC GTCCGCG CCATGGC CATCTAC AAGCAGT CACAGCA CATGACG GAGGTTG TGAGGCG
          GCGGGCC GTGGGCG CAGGCGC GGTACCG GTAGATG TTCGTCA GTGTCGT GTACTGC CTCCAAC ACTCCGC
                                        p53
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
          .CysPro HisHisGlu ArgCys SerAsp SerAspGly LeuAla ProPro GlnHisLeu IleArg ValGlu
   3711   CTGCCCC CACCATG AGCGCTG CTCAGAT AGCGATG GTCTGGC CCCTCCT CAGCATC TTATCCG AGTGGAA
          GACGGGG GTGGTAC TCGCGAC GAGTCTA TCGCTAC CAGACCG GGGAGGA GTCGTAG AATAGGC TCACCTT
```

```
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                GlyAsnLeu ArgVal GluTyr LeuAspAsp ArgAsn ThrPhe ArgHisSer ValVal ValPro TyrGluPro·
          3781  GGAAATT TGCGTGT GGAGTAT TTGGATG ACAGAAA CACTTTT CGACATA GTGTGGT GGTGCCC TATGAGC
          5     CCTTTAA ACGCACA CCTCATA AACCTAC TGTCTTT GTGAAAA GCTGTAT CACACCA CCACGGG ATACTCG
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                ..ProGlu ValGly SerAspCys ThrThr IleHis TyrAsnTyr MetCys AsnSer SerCysMet GlyGly·
          3851  CGCCTGA GGTTGGC TCTGACT GTACCAC CATCCAC TACAACT ACATGTG TAACAGT TCCTGCA TGGGCGG
          10    GCGGACT CCAACCG AGACTGA CATGGTG GTAGGTG ATGTTGA TGTACAC ATTGTCA AGGACGT ACCCGCC
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                .MetAsn ArgArgPro IleLeu ThrIle IleThrLeu GluAsp SerSer GlyAsnLeu LeuGly ArgAsn
          3921  CATGAAC CGGAGGC CCATCCT CACCATC ATCACAC TGGAAGA CTCCAGT GGTAATC TACTGGG ACGGAAC
          15    GTACTTG GCCTCCG GGTAGGA GTGGTAG TAGTGTG ACCTTCT GAGGTCA CCATTAG ATGACCC TGCCTTG
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                SerPheGlu ValArg ValCys AlaCysPro GlyArg AspArg ArgThrGlu GluGlu AsnLeu ArgLysLys·
          3991  AGCTTTG AGGTGCG TGTTTGT GCCTGTC CTGGGAG AGACCGG CGCACAG AGGAAGA GAATCTC CGCAAGA
          20    TCGAAAC TCCACGC ACAAACA CGGACAG GACCCTC TCTGGCC GCGTGTC TCCTTCT CTTAGAG GCGTTCT
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                ..GlyGlu ProHis HisGluLeu ProPro GlySer ThrLysArg AlaLeu ProAsn AsnThrSer SerSer·
          4061  AAGGGGA GCCTCAC CACGAGC TGCCCCC AGGGAGC ACTAAGC GAGCACT GCCCAAC AACACCA GCTCCTC
          25    TTCCCCT CGGAGTG GTGCTCG ACGGGGG TCCCTCG TGATTCG CTCGTGA CGGGTTG TTGTGGT CGAGGAG
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                .ProGln ProLysLys LysPro LeuAsp GlyGluTyr PheThr LeuGln IleArgGly ArgGlu ArgPhe
          4131  TCCCCAG CCAAAGA AGAAACC ACTGGAT GGAGAAT ATTTCAC CCTTCAG ATCCGTG GGCGTGA GCGCTTC
          30    AGGGGTC GGTTTCT TCTTTGG TGACCTA CCTCTTA TAAAGTG GGAAGTC TAGGCAC CCGCACT CGCGAAG
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                GluMetPhe ArgGlu LeuAsn GluAlaLeu GluLeu LysAsp AlaGlnAla GlyLys GluPro GlyGlySer·
          4201  GAGATGT TCCGAGA GCTGAAT GAGGCCT TGGAACT CAAGGAT GCCCAGG CTGGGAA GGAGCCA GGGGGGA
          35    CTCTACA AGGCTCT CGACTTA CTCCGGA ACCTTGA GTTCCTA CGGGTCC GACCCTT CCTCGGT CCCCCCT
                                        p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                ..ArgAla HisSer SerHisLeu LysSer LysLys GlyGlnSer ThrSer ArgHis LysLysLeu MetPhe·
          4271  GCAGGGC TCACTCC AGCCACC TGAAGTC CAAAAAG GGTCAGT CTACCTC CCGCCAT AAAAAAC TCATGTT
          40    CGTCCCG AGTGAGG TCGGTGG ACTTCAG GTTTTTC CCAGTCA GATGGAG GGCGGTA TTTTTTG AGTACAA
                      p53
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                .LysThr GluGlyPro AspSer Asp***
          4341  CAAGACA GAAGGGC CTGACTC AGACTGA ACGCGTT TTTTATC CCGGGCT CGAGGGT ACCGGAT CCTTTTT
          45    GTTCTGT CTTCCCG GACTGAG TCTGACT TGCGCAA AAAATAG GGCCCGA GCTCCCA TGGCCTA GGAAAAA
          4411  ATAGCTA ATTAGTC ACGTACC TTTGAGA GTACCAC TTCAGCT ACCTCTT TTGTGTC TCAGAGT AACTTTC
                TATCGAT TAATCAG TGCATGG AAACTCT CATGGTG AAGTCGA TGGAGAA AACACAG AGTCTCA TTGAAAG
                ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                                Right Arm
          50    4481  TTTAATC AATTCCA AAACAGT ATATGAT TTTCCAT TTCTTTC AAAGATG TAGTTTA CATCTGC TCCTTTG
                      AAATTAG TTAAGGT TTTGTCA TATACTA AAAGGTA AAGAAAG TTTCTAC ATCAAAT GTAGACG AGGAAAC
                                                Right Arm
                4551  TTGAAAA GTAGCCT GAGCACT TCTTTTC TACCATG AATTACA GCTGGCA AGATCAA TTTTTCC CAGTTCT
                      AACTTTT CATCGGA CTCGTGA AGAAAAG ATGGTAC TTAATGT CGACCGT TCTAGTT AAAAAGG GTCAAGA
          55                                    Right Arm
                4621  GGACATT TTATTTT TTTTAAG TAGTGTG CTACATA TTTCAAT ATTTCCA GATTGTA CAGCGAT CATTAAA
                      CCTGTAA AATAAAA AAAATTC ATCACAC GATGTAT AAAGTTA TAAAGGT CTAACAT GTCGCTA GTAATTT
                                                Right Arm
                4691  GGAGTAC GTCCCAT GTTATCC AGCAAGT CAGTATC AGCACCT TTGTTCA ATAGAAG TTTAACC ATTGTTA
          60    CCTCATG CAGGGTA CAATAGG TCGTTCA GTCATAG TCGTGGA AACAAGT TATCTTC AAATTGG TAACAAT
                                                Right Arm
                4761  AATTTTT ATTTGAT ACGGCTA TATGTAG AGGAGTT AACCGAT CCGTGTT TGAAATA TCTACAT CCGCCGA
                      TTAAAAA TAAACTA TGCCGAT ATACATC TCCTCAA TTGGCTA GGCACAA ACTTTAT AGATGTA GGCGGCT
                                                Right Arm
          65    4831  ATGAGCC AATAGAA GTTTAAC CAAATTA ACTTTGT TAAGGTA AGCTGCC AAACACA AAGGAGT AAAGCCT
                      TACTCGG TTATCTT CAAATTG GTTTAAT TGAAACA ATTCCAT TCGACGG TTTGTGT TTCCTCA TTTCGGA
                                                Right Arm
                4901  CCGCTGT AAAGAAC ATTGTTT ACATAGT TATTCTT CAACAGA TCTTTCA CTATTTT GTAGTCG TCTCTCA
                      GGCGACA TTTCTTG TAACAAA TGTATCA ATAAGAA GTTGTCT AGAAAGT GATAAAA CATCAGC AGAGAGT
          70                                    Right Arm
                4971  ACACCGC ATCATGC AGACAAG AAGTTGT GCATTCA GTAACTA CAGGTTT AGCTCCA TACCTCA TCAAGAT
                      TGTGGCG TAGTACG TCTGTTC TTCAACA CGTAAGT CATTGAT GTCCAAA TCGAGGT ATGGAGT AGTTCTA
                                                Right Arm
                5041  TTTTATA GCCTCGG TATTCTT GAACATT ACAGCCA TTTCAAG AGGAGAT TGTAGAG TACCATA TTCCGTG
```

```
          AAAATAT CGGAGCC ATAAGAA CTTGTAA TGTCGGT AAAGTTC TCCTCTA ACATCTC ATGGTAT AAGGCAC
                                  Right Arm
5111      TTAGGGT CGAATCC ATTGTCC AAAAACC TATTTAG AGATGCA TTGTCAT TATCCAT GATAGCC TCACAGA
          AATCCCA GCTTAGG TAACAGG TTTTTGG ATAAATC TCTACGT AACAGTA ATAGGTA CTATCGG AGTGTCT
                                  Right Arm
5181      CGTATAT GTAAGCC ATCTTGA ATGTATA ATTTTGT TGTTTTC AACAACC GCTCGTG AACAGCT TCTATAC
          GCATATA CATTCGG TAGAACT TACATAT TAAAACA ACAAAAG TTGTTGG CGAGCAC TTGTCGA AGATATG
                                  Right Arm
5251      TTTTTCA TTTTCTT CATGATT AATATAG TTTACGG AATATAA GTATACA AAAAGTT TATAGTA ATCTCAT
          AAAAAGT AAAAGAA GTACTAA TTATATC AAATGCC TTATATT CATATGT TTTTCAA ATATCAT TAGAGTA
                                  Right Arm
5321      AATATCT GAAACAC ATACATA AAACATG GAAGAAT TACACGA TGTCGTT GAGATAA ATGGCTT TTTATTG
          TTATAGA CTTTGTG TATGTAT TTTGTAC CTTCTTA ATGTGCT ACAGCAA CTCTATT TACCGAA AAATAAC
                                  Right Arm
5391      TCATAGT TTACAAA TTCGCAG TAATCTT CATCTTT TACGAAT ATTGCAG AATCTGT TTTATCC AACCAGT
          AGTATCA AATGTTT AAGCGTC ATTAGAA GTAGAAA ATGCTTA TAACGTC TTAGACA AAATAGG TTGGTCA
                                  Right Arm
5461      GATTTTT GTATAAT ATAACTG GTATCCT ATCTTCC GATAGAA TGCTGTT ATTTAAC ATTTTTG CACCTAT
          CTAAAAA CATATTA TATTGAC CATAGGA TAGAAGG CTATCTT ACGACAA TAAATTG TAAAAAC GTGGATA
                                  Right Arm
5531      TAAGTTA CATCTGT CAAATCC ATCTTTC CAACTGA CTTTATG TAACGAT GCGAAAT AGCATTT ATCACTA
          ATTCAAT GTAGACA GTTTAGG TAGAAAG GTTGACT GAAATAC ATTGCTA CGCTTTA TCGTAAA TAGTGAT
                                  Right Arm
5601      TGTCGTA CCCAATT ATCATGA CAAGATT CTCTTAA ATACGTA ATCTTAT TATCTCT TGCATAT TCGTAAT
          ACAGCAT GGGTTAA TAGTACT GTTCTAA GAGAATT TATGCAT TAGAATA ATAGAGA ACGTATA AGCATTA
                                  Right Arm
5671      AGTAATT GTAAAGA GTATACG ATAACAG TATAGAT ATACACG TGATATA AATATTT AACCCCA TTCCTGA
          TCATTAA CATTTCT CATATGC TATTGTC ATATCTA TATGTGC ACTATAT TTATAAA TTGGGGT AAGGACT
                                  Right Arm
5741      GTAAAAT AATTACG ATATTAC ATTTCCT TTTATTA TTTTTAT GTTTTAG TTATTTG TTAGGTT ATACAAA
          CATTTTA TTAATGC TATAATG TAAAGGA AAATAAT AAAAATA CAAAATC AATAAAC AATCCAA TATGTTT
                                  Right Arm
5811      AATTATG TTTATTT GTGTATA TTTAAAG CGTCGTT AAGAATA AGCTTAG TTAACAT ATTATCG CTTAGGT
          TTAATAC AAATAAA CACATAT AAATTTC GCAGCAA TTCTTAT TCGAATC AATTGTA TAATAGC GAATCCA
                                  Right Arm
5881      TTTGTAG TATTTGA ATCCTTT CTTTAAA TGGATTA TTTTTCC AATGCAT ATTTATA GCTTCAT CCAAAGT
          AAACATC ATAAACT TAGGAAA GAAATTT ACCTAAT AAAAAGG TTACGTA TAAATAT CGAAGTA GGTTTCA
                                  Right Arm
5951      ATAACAT TTAACAT TCAGAAT TGCGGCC GCAATTC AATTCGT AATCATG GTCATAG CTGTTTC CTGTGTG
          TATTGTA AATTGTA AGTCTTA ACGCCGG CGTTAAG TTAAGCA TTAGTAC CAGTATC GACAAAG GACACAC
                                  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Right Arm
6021      AAATTGT TATCCGC TCACAAT TCCACAC AACATAC GAGCCGG AAGCATA AAGTGTA AAGCCTG GGGTGCC
          TTTAACA ATAGGCG AGTGTTA AGGTGTG TTGTATG CTCGGCC TTCGTAT TTCACAT TTCGGAC CCCACGG
6091      TAATGAG TGAGCTA ACTCACA TTAATTG CGTTGCG CTCACTG CCCGCTT TCCAGTC GGGAAAC CTGTCGT
          ATTACTC ACTCGAT TGAGTGT AATTAAC GCAACGC GAGTGAC GGGCGAA AGGTCAG CCCTTTG GACAGCA
6161      GCCAGCT GCATTAA TGAATCG GCCAACG CGCGGGG AGAGGCG GTTTGCG TATTGGG CGCTCTT CCGCTTC
          CGGTCGA CGTAATT ACTTAGC CGGTTGC GCGCCCC TCTCCGC CAAACGC ATAACCC GCGAGAA GGCGAAG
6231      CTCGCTC ACTGACT CGCTGCG CTCGGTC GTTCGGC TGCGGCG AGCGGTA TCAGCTC ACTCAAA GGCGGTA
          GAGCGAG TGACTGA GCGACGC GAGCCAG CAAGCCG ACGCCGC TCGCCAT AGTCGAG TGAGTTT CCGCCAT
6301      ATACGGT TATCCAC AGAATCA GGGGATA ACGCAGG AAAGAAC ATGTGAG CAAAAGG CCAGCAA AAGGCCA
          TATGCCA ATAGGTG TCTTAGT CCCCTAT TGCGTCC TTTCTTG TACACTC GTTTTCC GGTCGTT TTCCGGT
6371      GGAACCG TAAAAAG GCCGCGT TGCTGGC GTTTTTC CATAGGC TCCGCCC CCCTGAC GAGCATC ACAAAAA
          CCTTGGC ATTTTTC CGGCGCA ACGACCG CAAAAAG GTATCCG AGGCGGG GGGACTG CTCGTAG TGTTTTT
6441      TCGACGC TCAAGTC AGAGGTG GCGAAAC CCGACAG GACTATA AAGATAC CAGGCGT TTCCCCC TGGAAGC
          AGCTGCG AGTTCAG TCTCCAC CGCTTTG GGCTGTC CTGATAT TTCTATG GTCCGCA AAGGGGG ACCTTCG
6511      TCCCTCG TGCGCTC TCCTGTT CCGACCC TGCCGCT TACCGGA TACCTGT CCGCCTT TCTCCCT TCGGGAA
          AGGGAGC ACGCGAG AGGACAA GGCTGGG ACGGCGA ATGGCCT ATGGACA GGCGGAA AGAGGGA AGCCCTT
6581      GCGTGGC GCTTTCT CATAGCT CACGCTG TAGGTAT CTCAGTT CGGTGTA GGTCGTT CGCTCCA AGCTGGG
          CGCACCG CGAAAGA GTATCGA GTGCGAC ATCCATA GAGTCAA GCCACAT CCAGCAA GCGAGGT TCGACCC
6651      CTGTGTG CACGAAC CCCCCGT TCAGCCC GACCGCT GCGCCTT ATCCGGT AACTATC GTCTTGA GTCCAAC
          GACACAC GTGCTTG GGGGGCA AGTCGGG CTGGCGA CGCGGAA TAGGCCA TTGATAG CAGAACT CAGGTTG
6721      CCGGTAA GACACGA CTTATCG CCACTGG CAGCAGC CACTGGT AACAGGA TTAGCAG AGCGAGG TATGTAG
          GGCCATT CTGTGCT GAATAGC GGTGACC GTCGTCG GTGACCA TTGTCCT AATCGTC TCGCTCC ATACATC
6791      GCGGTGC TACAGAG TTCTTGA AGTGGTG GCCTAAC TACGGCT ACACTAG AAGGACA GTATTTG GTATCTG
          CGCCACG ATGTCTC AAGAACT TCACCAC CGGATTG ATGCCGA TGTGATC TTCCTGT CATAAAC CATAGAC
6861      CGCTCTG CTGAAGC CAGTTAC CTTCGGA AAAAGAG TTGGTAG CTCTTGA TCCGGCA AACAAAC CACCGCT
          GCGAGAC GACTTCG GTCAATG GAAGCCT TTTTCTC AACCATC GAGAACT AGGCCGT TTGTTTG GTGGCGA
6931      GGTAGCG GTGGTTT TTTTGTT TGCAAGC AGCAGAT TACGCGC AGAAAAA AAGGATC TCAAGAA GATCCTT
          CCATCGC CACCAAA AAAACAA ACGTTCG TCGTCTA ATGCGCG TCTTTTT TTCCTAG AGTTCTT CTAGGAA
7001      TGATCTT TTCTACG GGGTCTG ACGCTCA GTGGAAC GAAAACT CACGTTA AGGGATT TTGGTCA TGAGATT
          ACTAGAA AAGATGC CCCAGAC TGCGAGT CACCTTG CTTTTGA GTGCAAT TCCCTAA AACCAGT ACTCTAA
7071      ATCAAAA AGGATCT TCACCTA GATCCTT TTAAATT AAAAATG AAGTTTT AAATCAA TCTAAAG TATATAT
          TAGTTTT TCCTAGA AGTGGAT CTAGGAA AATTTAA TTTTTAC TTCAAAA TTTAGTT AGATTTC ATATATA
```

```
7141    GAGTAAA CTTGGTC TGACAGT TACCAAT GCTTAAT CAGTGAG GCACCTA TCTCAGC GATCTGT CTATTTC
        CTCATTT GAACCAG ACTGTCA ATGGTTA CGAATTA GTCACTC CGTGGAT AGAGTCG CTAGACA GATAAAG
        ~~~~~~~~~~~~~~~~~~~~~~
                                              Amp resistance gene
7211    GTTCATC CATAGTT GCCTGAC TCCCCGT CGTGTAG ATAACTA CGATACG GGAGGGC TTACCAT CTGGCCC
        CAAGTAG GTATCAA CGGACTG AGGGGCA GCACATC TATTGAT GCTATGC CCTCCCG AATGGTA GACCGGG
                                   Amp resistance gene
7281    CAGTGCT GCAATGA TACCGCG AGACCCA CGCTCAC CGGCTCC AGATTTA TCAGCAA TAAACCA GCCAGCC
        GTCACGA CGTTACT ATGGCGC TCTGGGT GCGAGTG GCCGAGG TCTAAAT AGTCGTT ATTTGGT CGGTCGG
                                   Amp resistance gene
7351    GGAAGGG CCGAGCG CAGAAGT GGTCCTG CAACTTT ATCCGCC TCCATCC AGTCTAT TAATTGT TGCCGGG
        CCTTCCC GGCTCGC GTCTTCA CCAGGAC GTTGAAA TAGGCGG AGGTAGG TCAGATA ATTAACA ACGGCCC
                                   Amp resistance gene
7421    AAGCTAG AGTAAGT AGTTCGC CAGTTAA TAGTTTG CGCAACG TTGTTGC CATTGCT ACAGGCA TCGTGGT
        TTCGATC TCATTCA TCAAGCG GTCAATT ATCAAAC GCGTTGC AACAACG GTAACGA TGTCCGT AGCACCA
                                   Amp resistance gene
7491    GTCACGC TCGTCGT TTGGTAT GGCTTCA TTCAGCT CCGGTTC CCAACGA TCAAGGC GAGTTAC ATGATCC
        CAGTGCG AGCAGCA AACCATA CCGAAGT AAGTCGA GGCCAAG GGTTGCT AGTTCCG CTCAATG TACTAGG
                                   Amp resistance gene
7561    CCCATGT TGTGCAA AAAAGCG GTTAGCT CCTTCGG TCCTCCG ATCGTTG TCAGAAG TAAGTTG GCCGCAG
        GGGTACA ACACGTT TTTTCGC CAATCGA GGAAGCC AGGAGGC TAGCAAC AGTCTTC ATTCAAC CGGCGTC
                                   Amp resistance gene
7631    TGTTATC ACTCATG GTTATGG CAGCACT GCATAAT TCTCTTA CTGTCAT GCCATCC GTAAGAT GCTTTTC
        ACAATAG TGAGTAC CAATACC GTCGTGA CGTATTA AGAGAAT GACAGTA CGGTAGG CATTCTA CGAAAAG
                                   Amp resistance gene
7701    TGTGACT GGTGAGT ACTCAAC CAAGTCA TTCTGAG AATAGTG TATGCGG CGACCGA GTTGCTC TTGCCCG
        ACACTGA CCACTCA TGAGTTG GTTCAGT AAGACTC TTATCAC ATACGCC GCTGGCT CAACGAG AACGGGC
                                   Amp resistance gene
7771    GCGTCAA TACGGGA TAATACC GCGCCAC ATAGCAG AACTTTA AAAGTGC TCATCAT TGGAAAA CGTTCTT
        CGCAGTT ATGCCCT ATTATGG CGCGGTG TATCGTC TTGAAAT TTTCACG AGTAGTA ACCTTTT GCAAGAA
                                   Amp resistance gene
7841    CGGGGCG AAAACTC TCAAGGA TCTTACC GCTGTTG AGATCCA GTTCGAT GTAACCC ACTCGTG CACCCAA
        GCCCCGC TTTTGAG AGTTCCT AGAATGG CGACAAC TCTAGGT CAAGCTA CATTGGG TGAGCAC GTGGGTT
                                   Amp resistance gene
7911    CTGATCT TCAGCAT CTTTTAC TTTCACC AGCGTTT CTGGGTG AGCAAAA ACAGGAA GGCAAAA TGCCGCA
        GACTAGA AGTCGTA GAAAATG AAAGTGG TCGCAAA GACCCAC TCGTTTT TGTCCTT CCGTTTT ACGGCGT
                                   Amp resistance gene
7981    AAAAAGG GAATAAG GGCGACA CGGAAAT GTTGAAT ACTCATA CTCTTCC TTTTTCA ATATTAT TGAAGCA
        TTTTTCC CTTATTC CCGCTGT GCCTTTA CAACTTA TGAGTAT GAGAAGG AAAAAGT TATAATA ACTTCGT
                                                        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                Amp resistance gene
8051    TTTATCA GGGTTAT TGTCTCA TGAGCGG ATACATA TTTGAAT GTATTTA GAAAAAT AAACAAA TAGGGGT
        AAATAGT CCCAATA ACAGAGT ACTCGCC TATGTAT AAACTTA CATAAAT CTTTTTA TTTGTTT ATCCCCA
8121    TCCGCGC ACATTTC CCCGAAA AGTGCCA CCTGACG TCTAAGA AACCATT ATTATCA TGACATT AACCTAT
        AGGCGCG TGTAAAG GGGCTTT TCACGGT GGACTGC AGATTCT TTGGTAA TAATAGT ACTGTAA TTGGATA
8191    AAAAATA GGCGTAT CACGAG
       TTTTTAT CCGCATA GTGCTC
```

FIGURE 2A

```
                    1                                                  50
        mCEA(6D)    ATGGAGTCTC CCTCGGCCCC TCCCCACAGA TGGTGCATCC CCTGGCAGAG
mCEA(6D,1st&2nd)    ATGGAGTCTC CCTCGGCCCC TCCCCACAGA TGGTGCATCC CCTGGCAGAG

                   51                                                 100
        mCEA(6D)    GCTCCTGCTC ACAGCCTCAC TTCTAACCTT CTGGAACCCG CCCACCACTG
mCEA(6D,1st&2nd)    GCTCCTGCTC ACAGCCTCAC TTCTAACCTT CTGGAACCCG CCCACCACTG

                  101                                                 150
        mCEA(6D)    CCAAGCTCAC TATTGAATCC ACGCCGTTCA ATGTCGCAGA GGGGAAGGAG
mCEA(6D,1st&2nd)    CCAAGCTCAC TATTGAATCC ACGCCGTTCA ATGTCGCAGA GGGGAAGGAG

                  151                                                 200
        mCEA(6D)    GTGCTTCTAC TTGTCCACAA TCTGCCCCAG CATCTTTTTG GCTACAGCTG
mCEA(6D,1st&2nd)    GTGCTTCTAC TTGTCCACAA TCTGCCCCAG CATCTTTTTG GCTACAGCTG

                  201                                                 250
        mCEA(6D)    GTACAAAGGT GAAAGAGTGG ATGGCAACCG TCAAATTATA GGATATGTAA
mCEA(6D,1st&2nd)    GTACAAAGGT GAAAGAGTGG ATGGCAACCG TCAAATTATA GGATATGTAA

                  251                                                 300
        mCEA(6D)    TAGGAACTCA ACAAGCTACC CCAGGGCCCG CATACAGTGG TCGAGAGATA
mCEA(6D,1st&2nd)    TAGGAACTCA ACAAGCTACC CCAGGGCCCG CATACAGTGG TCGAGAGATA

                  301                                                 350
        mCEA(6D)    ATATACCCCA ATGCATCCCT GCTGATCCAG AACATCATCC AGAATGACAC
mCEA(6D,1st&2nd)    ATATACCCCA ATGCATCCCT GCTGATCCAG AACATCATCC AGAATGACAC

                  351                                                 400
        mCEA(6D)    AGGATTCTAC ACCCTACACG TCATAAAGTC AGATCTTGTG AATGAAGAAG
mCEA(6D,1st&2nd)    AGGATTCTAC ACCCTACACG TCATAAAGTC AGATCTTGTG AATGAAGAAG

                  401                                                 450
        mCEA(6D)    CAACTGGCCA GTTCCGGGTA TACCCGGAGC TGCCCAAGCC CTCCATCTCC
mCEA(6D,1st&2nd)    CAACTGGCCA GTTCCGGGTA TACCCGGAAC TCCCTAAGCC TTCTATTAGC

                  451                                                 500
        mCEA(6D)    AGCAACAACT CCAAACCCGT GGAGGACAAG GATGCTGTGG CCTTCACCTG
mCEA(6D,1st&2nd)    TCCAATAATA GTAAGCCTGT CGAAGACAAA GATGCCGTCG CTTTTACATG

                  501                                                 550
        mCEA(6D)    TGAACCTGAG ACTCAGGACG CAACCTACCT GTGGTGGGTA AACAATCAGA
mCEA(6D,1st&2nd)    CGAGCCCGAA ACTCAAGACG CAACATATCT CTGGTGGGTG AACAACCAGT

                  551                                                 600
        mCEA(6D)    GCCTCCCGGT CAGTCCCAGG CTGCAGCTGT CCAATGGCAA CAGGACCCTC
mCEA(6D,1st&2nd)    CCCTGCCTGT GTCCCCTAGA CTCCAACTCA GCAACGGAAA TAGAACTCTG

                  601                                                 650
        mCEA(6D)    ACTCTATTCA ATGTCACAAG AAATGACACA GCAAGCTACA AATGTGAAAC
mCEA(6D,1st&2nd)    ACCCTGTTTA ACGTGACCAG GAACGACACA GCAAGCTACA AATGCGAAAC
```

FIGURE 2B

```
                    651                                                 700
          mCEA(6D)  CCAGAACCCA GTGAGTGCCA GGCGCAGTGA TTCAGTCATC CTGAATGTCC
mCEA(6D,1st&2nd)    CCAAAATCCA GTCAGCGCCA GGAGGTCTGA TTCAGTGATT CTCAACGTGC

                    701                                                 750
          mCEA(6D)  TCTATGGCCC GGATGCCCCC ACCATTTCCC CTCTAAACAC ATCTTACAGA
mCEA(6D,1st&2nd)    TTTACGGACC CGATGCTCCT ACAATCAGCC CTCTAAACAC AAGCTATAGA

                    751                                                 800
          mCEA(6D)  TCAGGGGAAA ATCTGAACCT CTCCTGCCAC GCAGCCTCTA ACCCACCTGC
mCEA(6D,1st&2nd)    TCAGGGGAAA ATCTGAATCT GAGCTGTCAT GCCGCTAGCA ATCCTCCCGC

                    801                                                 850
          mCEA(6D)  ACAGTACTCT TGGTTTGTCA ATGGGACTTT CCAGCAATCC ACCCAAGAGC
mCEA(6D,1st&2nd)    CCAATACAGC TGGTTTGTCA ATGGCACTTT CCAACAGTCC ACCCAGGAAC

                    851                                                 900
          mCEA(6D)  TCTTTATCCC CAACATCACT GTGAATAATA GTGGATCCTA TACGTGCCAA
mCEA(6D,1st&2nd)    TGTTCATTCC CAATATTACC GTGAACAATA GTGGATCCTA CACGTGCCAA

                    901                                                 950
          mCEA(6D)  GCCCATAACT CAGACACTGG CCTCAATAGG ACCACAGTCA CGACGATCAC
mCEA(6D,1st&2nd)    GCTCACAATA GCGACACCGG ACTCAACCGC ACAACCGTGA CGACGATTAC

                    951                                                1000
          mCEA(6D)  AGTCTATGAG CCACCCAAAC CCTTCATCAC CAGCAACAAC TCCAACCCCG
mCEA(6D,1st&2nd)    CGTGTATGAG CCACCAAAAC CATTCATAAC TAGTAACAAT TCTAACCCAG

                   1001                                                1050
          mCEA(6D)  TGGAGGATGA GGATGCTGTA GCCTTAACCT GTGAACCTGA GATTCAGAAC
mCEA(6D,1st&2nd)    TTGAGGATGA GGACGCAGTT GCATTAACTT GTGAGCCAGA GATTCAAAAT

                   1051                                                1100
          mCEA(6D)  ACAACCTACC TGTGGTGGGT AAATAATCAG AGCCTCCCGG TCAGTCCCAG
mCEA(6D,1st&2nd)    ACCACTTATT TATGGTGGGT CAATAACCAA AGTTTGCCGG TTAGCCCACG

                   1101                                                1150
          mCEA(6D)  GCTGCAGCTG TCCAATGACA ACAGGACCCT CACTCTACTC AGTGTCACAA
mCEA(6D,1st&2nd)    CTTGCAGTTG TCTAATGATA ACCGCACATT GACACTCCTG TCCGTTACTC

                   1151                                                1200
          mCEA(6D)  GGAATGATGT AGGACCCTAT GAGTGTGGAA TCCAGAACGA ATTAAGTGTT
mCEA(6D,1st&2nd)    GCAATGATGT AGGACCTTAT GAGTGTGGCA TTCAGAATGA ATTATCCGTT

                   1201                                                1250
          mCEA(6D)  GACCACAGCG ACCCAGTCAT CCTGAATGTC CTCTATGGCC CAGACGACCC
mCEA(6D,1st&2nd)    GATCACTCCG ACCCTGTTAT CCTTAATGTT TTGTATGGCC CAGACGACCC

                   1251                                                1300
          mCEA(6D)  CACCATTTCC CCCTCATACA CCTATTACCG TCCAGGGGTG AACCTCAGCC
mCEA(6D,1st&2nd)    AACTATATCT CCATCATACA CCTACTACCG TCCCGGCGTG AACTTGAGCC
```

FIGURE 2C

```
                    1301                                                1350
        mCEA(6D)    TCTCCTGCCA TGCAGCCTCT AACCCACCTG CACAGTATTC TTGGCTGATT
mCEA(6D,1st&2nd)    TTTCTTGCCA TGCAGCATCC AACCCCCCTG CACAGTACTC CTGGCTGATT

                    1351                                                1400
        mCEA(6D)    GATGGGAACA TCCAGCAACA CACACAAGAG CTCTTTATCT CCAACATCAC
mCEA(6D,1st&2nd)    GATGGAAACA TTCAGCAGCA TACTCAAGAG TTATTTATAA GCAACATAAC

                    1401                                                1450
        mCEA(6D)    TGAGAAGAAC AGCGGACTCT ATACCTGCCA GGCCAATAAC TCAGCCAGTG
mCEA(6D,1st&2nd)    TGAGAAGAAC AGCGGACTCT ATACTTGCCA GGCCAATAAC TCAGCCAGTG

                    1451                                                1500
        mCEA(6D)    GCCACAGCAG GACTACAGTC AAGACAATCA CAGTCTCTGC GGAGCTGCCC
mCEA(6D,1st&2nd)    GTCACAGCAG GACTACAGTT AAAACAATAA CTGTTTCCGC GGAGCTGCCC

                    1501                                                1550
        mCEA(6D)    AAGCCCTCCA TCTCCAGCAA CAACTCCAAA CCCGTGGAGG ACAAGGATGC
mCEA(6D,1st&2nd)    AAGCCCTCCA TCTCCAGCAA CAACTCCAAA CCCGTGGAGG ACAAGGATGC

                    1551                                                1600
        mCEA(6D)    TGTGGCCTTC ACCTGTGAAC CTGAGGCTCA GAACACAACC TACCTGTGGT
mCEA(6D,1st&2nd)    TGTGGCCTTC ACCTGTGAAC CTGAGGCTCA GAACACAACC TACCTGTGGT

                    1601                                                1650
        mCEA(6D)    GGGTAAATGG TCAGAGCCTC CCAGTCAGTC CCAGGCTGCA GCTGTCCAAT
mCEA(6D,1st&2nd)    GGGTAAATGG TCAGAGCCTC CCAGTCAGTC CCAGGCTGCA GCTGTCCAAT

                    1651                                                1700
        mCEA(6D)    GGCAACAGGA CCCTCACTCT ATTCAATGTC ACAAGAAATG ACGCAAGAGC
mCEA(6D,1st&2nd)    GGCAACAGGA CCCTCACTCT ATTCAATGTC ACAAGAAATG ACGCAAGAGC

                    1701                                                1750
        mCEA(6D)    CTATGTATGT GGAATCCAGA ACTCAGTGAG TGCAAACCGC AGTGACCCAG
mCEA(6D,1st&2nd)    CTATGTATGT GGAATCCAGA ACTCAGTGAG TGCAAACCGC AGTGACCCAG

                    1751                                                1800
        mCEA(6D)    TCACCCTGGA TGTCCTCTAT GGGCCGGACA CCCCCATCAT TTCCCCCCCA
mCEA(6D,1st&2nd)    TCACCCTGGA TGTCCTCTAT GGGCCGGACA CCCCCATCAT TTCCCCCCCA

                    1801                                                1850
        mCEA(6D)    GACTCGTCTT ACCTTTCGGG AGCGGACCTC AACCTCTCCT GCCACTCGGC
mCEA(6D,1st&2nd)    GACTCGTCTT ACCTTTCGGG AGCGGACCTC AACCTCTCCT GCCACTCGGC

                    1851                                                1900
        mCEA(6D)    CTCTAACCCA TCCCCGCAGT ATTCTTGGCG TATCAATGGG ATACCGCAGC
mCEA(6D,1st&2nd)    CTCTAACCCA TCCCCGCAGT ATTCTTGGCG TATCAATGGG ATACCGCAGC

                    1901                                                1950
        mCEA(6D)    AACACACACA AGTTCTCTTT ATCGCCAAAA TCACGCCAAA TAATAACGGG
mCEA(6D,1st&2nd)    AACACACACA AGTTCTCTTT ATCGCCAAAA TCACGCCAAA TAATAACGGG
```

FIGURE 2D

```
                      1951                                                2000
          mCEA(6D)    ACCTATGCCT GTTTTGTCTC TAACTTGGCT ACTGGCCGCA ATAATTCCAT
mCEA(6D,1st&2nd)      ACCTATGCCT GTTTTGTCTC TAACTTGGCT ACTGGCCGCA ATAATTCCAT

                      2001                                                2050
          mCEA(6D)    AGTCAAGAGC ATCACAGTCT CTGCATCTGG AACTTCTCCT GGTCTCTCAG
mCEA(6D,1st&2nd)      AGTCAAGAGC ATCACAGTCT CTGCATCTGG AACTTCTCCT GGTCTCTCAG

                      2051                                                2100
          mCEA(6D)    CTGGGGCCAC TGTCGGCATC ATGATTGGAG TGCTGGTTGG GGTTGCTCTG
mCEA(6D,1st&2nd)      CTGGGGCCAC TGTCGGCATC ATGATTGGAG TGCTGGTTGG GGTTGCTCTG

                      2101
          mCEA(6D)    ATATAG
mCEA(6D,1st&2nd)      ATATAG
```

FIGURE 3

A. Amino Acid Sequence Comparison of "Wild-Type KSA" (1) and Modified KSA (2)

```
1 MAPPQVLAFGLLLAAATATFAAAQEECVCENYKLAVNCFVNNNRQCQCTSVGAQNTVIC
2 MAPPQVLAFGLLLAAATATFAAAQEECVCENYKLAVNCFVNNNRQCQCTSVGAQNTVIC

1 SKLAAKCLVMKAEMNGSKLGRRAKPEGALQNNDGLYDPDCDESGLFKAKQCNGTSTCWC
2 SKLAAKCLVMKAEMNGSKLGRRAKPEGALQNNDGLYDPDCDESGLFKAKQCNGTSTCWC

1 VNTAGVRRTDKDTEITCSERVRTYWIIIELKHKAREKPYDSKSLRTALQKEITTRYQLD
2 VNTAGVRRTDKDTEITCSERVRTYWIIIELKHKAREKPYDSKSLRTALQKEITTRYQLD

1 PKFITSILYENNVITIDLVQNSSQKTQNDVDIADVAYYFEKDVKGESLFHSKKMDLTVN
2 PKFITSVLYENNVITIDLVQNSSQKTQNDVDIADVAYYFEKDVKGESLFHSKKMDLTVN

1 GEQLDLDPGQTLIYYVDEKAPEFSMQGLKAGVIAVIVVVVIAVVAGIVVLVISRKKRMA
2 GEQLDLDPGQTLIYYVDEKAPEFSMQGLKAGVIAVIVVVVIAVVAGIVVLVISRKKRMA

1 KYEKAEIKEMGEMHRELNA
2 KYEKAEIKEMGEMHRELNA
```

B. DNA Sequence of Modified KSA

```
atggcgcccccgcaggtcctcgcgttcgggcttctgcttgccgcggcgacggcgacttttgccgcagctcaggaa
gaatgtgtctgtgaaaactacaagctggccgtaaactgctttgtgaataataatcgtcaatgccagtgtacttca
gttggtgcacaaaatactgtcatttgctcaaagctggctgccaaatgtttggtgatgaaggcagaaatgaatggc
tcaaaacttgggagaagagcaaaacctgaaggggccctccagaacaatgatgggctttatgatcctgactgcgat
gagagcgggctctttaaggccaagcagtgcaacggcacctccacgtgctggtgtgtgaacactgctggggtcaga
agaacagacaaggacactgaaataacctgctctgagcgagtgagaacctactggatcatcattgaactaaaacac
aaagcaagagaaaaaccttatgatagtaaaagtttgcggactgcacttcagaaggagatcacaacgcgttatcaa
ctggatccaaaatttatcacgagtgtgttgtatgagaataatgttatcactattgatctggttcaaaattcttct
caaaaaactcagaatgatgtggacatagctgatgtggcttattattttgaaaaagatgttaaaggtgaatccttg
tttcattctaagaaaatggacctgacagtaaatggggaacaactggatctggatcctggtcaaactttaatttat
tatgttgatgaaaaagcacctgaattctcaatgcagggtctaaaagctggtgttattgctgttattgtggttgtg
gtgatagcagttgttgctggaattgttgtgctggttatttccagaaagaagagaatggcaaagtatgagaaggct
gagataaaggagatgggtgagatgcatagggaactcaatgcataa
```

FIGURE 4A

Construction of Modified KSA Plasmid

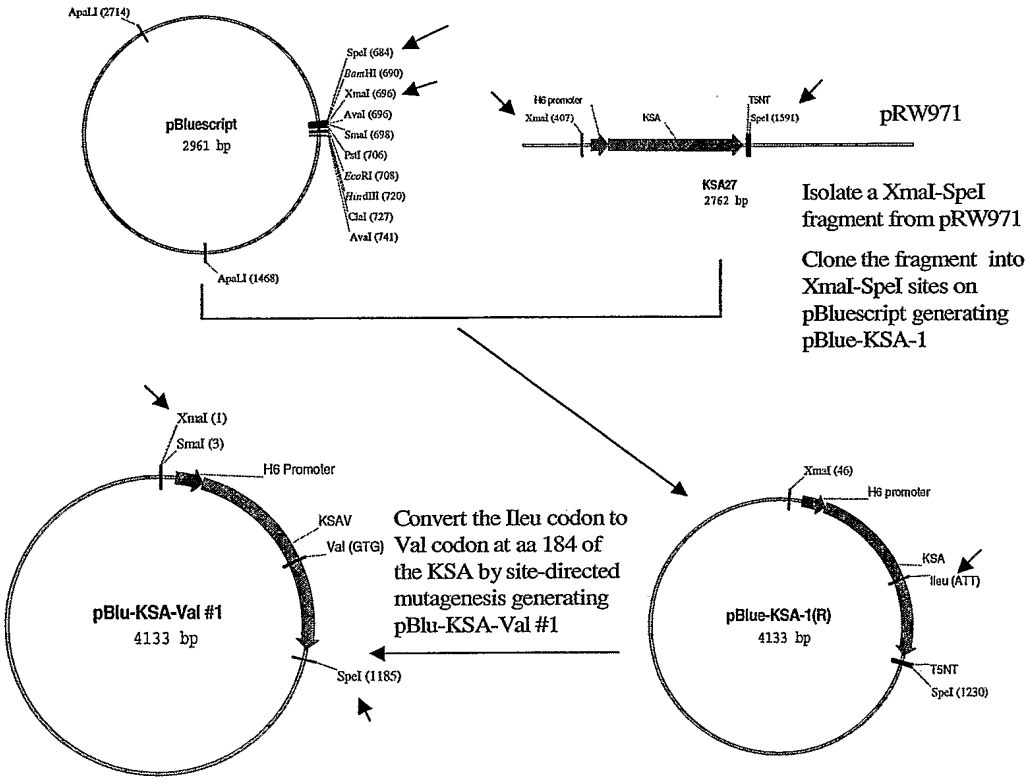


FIGURE 4B
Construction of Modified KSA Plasmid

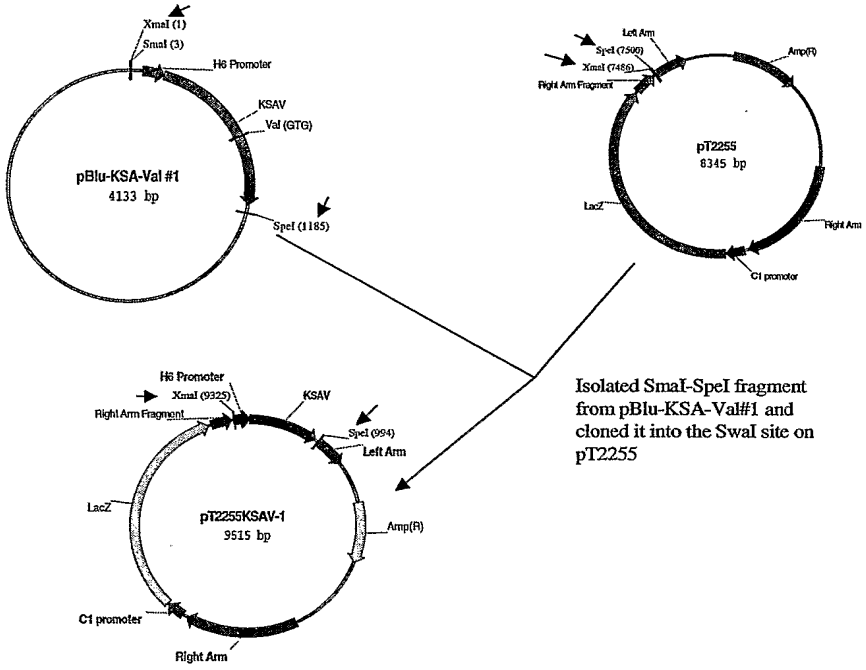
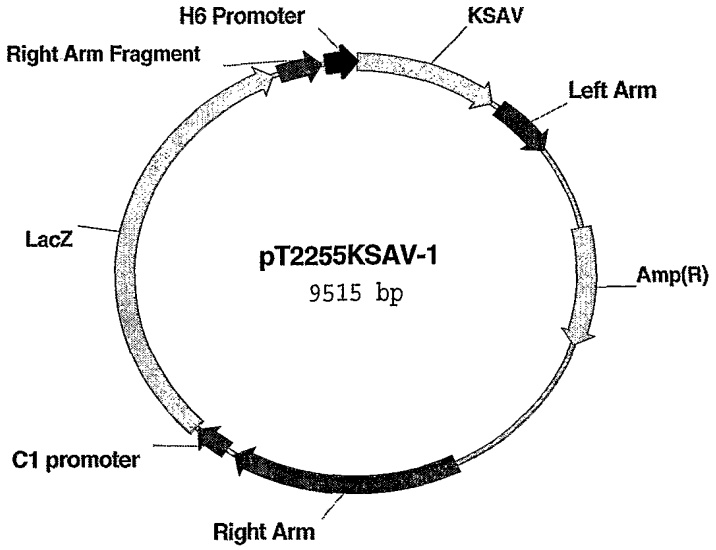


FIGURE 5

A. Plasmid Map of Modified KSA Expression Vector



B. DNA Sequence of Modified KSA Expression Vector

Promoter H6 for KSAV	**9930-9515**
KSAV	**1-945**
Left arm	**1002-1422**
Right arm	**4070-5590**
Right arm fragment	**9012-9299**

```
        MetAlaProPro GlnValLeu AlaPheGly LeuLeuLeuAla AlaAlaThr·
    1   ATGGCGCCCC CGCAGGTCCT CGCGTTCGGG CTTCTGCTTG CCGCGGCGAC
        TACCGCGGGG GCGTCCAGGA GCGCAAGCCC GAAGACGAAC GGCGCCGCTG
        .AlaThrPhe AlaAlaAlaGln GluGluCys ValCysGlu AsnTyrLysLeu·
   51   GGCGACTTTT GCCGCAGCTC AGGAAGAATG TGTCTGTGAA AACTACAAGC
        CCGCTGAAAA CGGCGTCGAG TCCTTCTTAC ACAGACACTT TTGATGTTCG
        ..AlaValAsn CysPheVal AsnAsnAsnArg GlnCysGln CysThrSer
  101   TGGCCGTAAA CTGCTTTGTG AATAATAATC GTCAATGCCA GTGTACTTCA
        ACCGGCATTT GACGAAACAC TTATTATTAG CAGTTACGGT CACATGAAGT
        ValGlyAlaGln AsnThrVal IleCysSer LysLeuAlaAla LysCysLeu·
  151   GTTGGTGCAC AAAATACTGT CATTTGCTCA AAGCTGGCTG CCAAATGTTT
        CAACCACGTG TTTTATGACA GTAAACGAGT TTCGACCGAC GGTTTACAAA
        .ValMetLys AlaGluMetAsn GlySerLys LeuGlyArg ArgAlaLysPro·
  201   GGTGATGAAG GCAGAAATGA ATGGCTCAAA ACTTGGGAGA AGAGCAAAAC
        CCACTACTTC CGTCTTTACT TACCGAGTTT TGAACCCTCT TCTCGTTTTG
        ..GluGlyAla LeuGlnAsn AsnAspGlyLeu TyrAspPro AspCysAsp
  251   CTGAAGGGGC CCTCCAGAAC AATGATGGGC TTTATGATCC TGACTGCGAT
        GACTTCCCCG GGAGGTCTTG TTACTACCCG AAATACTAGG ACTGACGCTA
        GluSerGlyLeu PheLysAla LysGlnCys AsnGlyThrSer ThrCysTrp·
  301   GAGAGCGGGC TCTTTAAGGC CAAGCAGTGC AACGGCACCT CCACGTGCTG
        CTCTCGCCCG AGAAATTCCG GTTCGTCACG TTGCCGTGGA GGTGCACGAC
        .CysValAsn ThrAlaGlyVal ArgArgThr AspLysAsp ThrGluIleThr·
  351   GTGTGTGAAC ACTGCTGGGG TCAGAAGAAC AGACAAGGAC ACTGAAATAA
        CACACACTTG TGACGACCCC AGTCTTCTTG TCTGTTCCTG TGACTTTATT
        ..CysSerGlu ArgValArg ThrTyrTrpIle IleIleGlu LeuLysHis
```

```
      401   CCTGCTCTGA GCGAGTGAGA ACCTACTGGA TCATCATTGA ACTAAAACAC
            GGACGAGACT CGCTCACTCT TGGATGACCT AGTAGTAACT TGATTTTGTG
            LysAlaArgGlu LysProTyr AspSerLys SerLeuArgThr AlaLeuGln·
      451   AAAGCAAGAG AAAAACCTTA TGATAGTAAA AGTTTGCGGA CTGCACTTCA
            TTTCGTTCTC TTTTTGGAAT ACTATCATTT TCAAACGCCT GACGTGAAGT
            .LysGluIle ThrThrArgTyr GlnLeuAsp ProLysPhe IleThrSerVal·
      501   GAAGGAGATC ACAACGCGTT ATCAACTGGA TCCAAAATTT ATCACGAGTG
            CTTCCTCTAG TGTTGCGCAA TAGTTGACCT AGGTTTTAAA TAGTGCTCAC
            ..LeuTyrGlu AsnAsnVal IleThrIleAsp LeuValGln AsnSerSer
      551   TGTTGTATGA GAATAATGTT ATCACTATTG ATCTGGTTCA AAATTCTTCT
            ACAACATACT CTTATTACAA TAGTGATAAC TAGACCAAGT TTTAAGAAGA
            GlnLysThrGln AsnAspVal AspIleAla AspValAlaTyr TyrPheGlu·
      601   CAAAAAACTC AGAATGATGT GGACATAGCT GATGTGGCTT ATTATTTTGA
            GTTTTTTGAG TCTTACTACA CCTGTATCGA CTACACCGAA TAATAAAACT
            .LysAspVal LysGlyGluSer LeuPheHis SerLysLys MetAspLeuThr·
      651   AAAAGATGTT AAAGGTGAAT CCTTGTTTCA TTCTAAGAAA ATGGACCTGA
            TTTTCTACAA TTTCCACTTA GGAACAAAGT AAGATTCTTT TACCTGGACT
            ..ValAsnGly GluGlnLeu AspLeuAspPro GlyGlnThr LeuIleTyr
      701   CAGTAAATGG GGAACAACTG GATCTGGATC CTGGTCAAAC TTTAATTTAT
            GTCATTTACC CCTTGTTGAC CTAGACCTAG GACCAGTTTG AAATTAAATA
            TyrValAspGlu LysAlaPro GluPheSer MetGlnGlyLeu LysAlaGly·
      751   TATGTTGATG AAAAAGCACC TGAATTCTCA ATGCAGGGTC TAAAAGCTGG
            ATACAACTAC TTTTTCGTGG ACTTAAGAGT TACGTCCCAG ATTTTCGACC
            .ValIleAla ValIleValVal ValValIle AlaValVal AlaGlyIleVal·
      801   TGTTATTGCT GTTATTGTGG TTGTGGTGAT AGCAGTTGTT GCTGGAATTG
            ACAATAACGA CAATAACACC AACACCACTA TCGTCAACAA CGACCTTAAC
            ..ValLeuVal IleSerArg LysLysArgMet AlaLysTyr GluLysAla
      851   TTGTGCTGGT TATTTCCAGA AAGAAGAGAA TGGCAAAGTA TGAGAAGGCT
            AACACGACCA ATAAAGGTCT TTCTTCTCTT ACCGTTTCAT ACTCTTCCGA
            GluIleLysGlu MetGlyGlu MetHisArg GluLeuAsnAla ***
      901   GAGATAAAGG AGATGGGTGA GATGCATAGG GAACTCAATG CATAAGAAGC
            CTCTATTTCC TCTACCCACT CTACGTATCC CTTGAGTTAC GTATTCTTCG
      951   TTATCGATAC CGTCGACCTC GAGGAATTCT TTTTATTGAT TAACTAGTTA
            AATAGCTATG GCAGCTGGAG CTCCTTAAGA AAAATAACTA ATTGATCAAT
     1001   ATCACGGCCG CTTATAAAGA TCTAAAATGC ATAATTTCTA AATAATGAAA
            TAGTGCCGGC GAATATTTCT AGATTTTACG TATTAAAGAT TTATTACTTT
     1051   AAAAAGTACA TCATGAGCAA CGCGTTAGTA TATTTTACAA TGGAGATTAA
            TTTTTCATGT AGTACTCGTT GCGCAATCAT ATAAAATGTT ACCTCTAATT
     1101   CGCTCTATAC CGTTCTATGT TTATTGATTC AGATGATGTT TTAGAAAAGA
            GCGAGATATG GCAAGATACA AATAACTAAG TCTACTACAA AATCTTTTCT
     1151   AAGTTATTGA ATATGAAAAC TTTAATGAAG ATGAAGATGA CGACGATGAT
            TTCAATAACT TATACTTTTG AAATTACTTC TACTTCTACT GCTGCTACTA
     1201   TATTGTTGTA AATCTGTTTT AGATGAAGAA GATGACGCGC TAAAGTATAC
            ATAACAACAT TTAGACAAAA TCTACTTCTT CTACTGCGCG ATTTCATATG
     1251   TATGGTTACA AAGTATAAGT CTATACTACT AATGGCGACT TGTGCAAGAA
            ATACCAATGT TTCATATTCA GATATGATGA TTACCGCTGA ACACGTTCTT
     1301   GGTATAGTAT AGTGAAAATG TTGTTAGATT ATGATTATGA AAAACCAAAT
            CCATATCATA TCACTTTTAC AACAATCTAA TACTAATACT TTTTGGTTTA
     1351   AAATCAGATC CATATCTAAA GGTATCTCCT TTGCACATAA TTTCATCTAT
            TTTAGTCTAG GTATAGATTT CCATAGAGGA AACGTGTATT AAAGTAGATA
     1401   TCCTAGTTTA GAATACCTGC AGCCAAGCTT GGCACTGGCC GTCGTTTTAC
            AGGATCAAAT CTTATGGACG TCGGTTCGAA CCGTGACCGG CAGCAAAATG
     1451   AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA TCGCCTTGCA
            TTGCAGCACT GACCCTTTTG GGACCGCAAT GGGTTGAATT AGCGGAACGT
     1501   GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA
            CGTGTAGGGG GAAAGCGGTC GACCGCATTA TCGCTTCTCC GGGCGTGGCT
     1551   TCGCCCTTCC CAACAGTTGC GCAGCCTGAA TGGCGAATGG CGCCTGATGC
            AGCGGGAAGG GTTGTCAACG CGTCGGACTT ACCGCTTACC GCGGACTACG
     1601   GGTATTTTCT CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATGGTGC
```

```
        CCATAAAAGA GGAATGCGTA GACACGCCAT AAAGTGTGGC GTATACCACG
1651    ACTCTCAGTA CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGCCCCGACA
        TGAGAGTCAT GTTAGACGAG ACTACGGCGT ATCAATTCGG TCGGGGCTGT
1701    CCCGCCAACA CCCGCTGACG CGCCCTGACG GGCTTGTCTG CTCCCGGCAT
        GGGCGGTTGT GGGCGACTGC GCGGGACTGC CCGAACAGAC GAGGGCCGTA
1751    CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT GTGTCAGAGG
        GGCGAATGTC TGTTCGACAC TGGCAGAGGC CCTCGACGTA CACAGTCTCC
1801    TTTTCACCGT CATCACCGAA ACGCGCGAGA CGAAAGGGCC TCGTGATACG
        AAAAGTGGCA GTAGTGGCTT TGCGCGCTCT GCTTTCCCGG AGCACTATGC
1851    CCTATTTTTA TAGGTTAATG TCATGATAAT AATGGTTTCT TAGACGTCAG
        GGATAAAAAT ATCCAATTAC AGTACTATTA TTACCAAAGA ATCTGCAGTC
1901    GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC
        CACCGTGAAA AGCCCCTTTA CACGCGCCTT GGGGATAAAC AAATAAAAAG
1951    TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT
        ATTTATGTAA GTTTATACAT AGGCGAGTAC TCTGTTATTG GGACTATTTA
2001    GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG
        CGAAGTTATT ATAACTTTTT CCTTCTCATA CTCATAAGTT GTAAAGGCAC
2051    TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC
        AGCGGGAATA AGGGAAAAAA CGCCGTAAAA CGGAAGGACA AAAACGAGTG
2101    CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG
        GGTCTTTGCG ACCACTTTCA TTTTCTACGA CTTCTAGTCA ACCCACGTGC
2151    AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
        TCACCCAATG TAGCTTGACC TAGAGTTGTC GCCATTCTAG GAACTCTCAA
2201    TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA
        AAGCGGGGCT TCTTGCAAAA GGTTACTACT CGTGAAAATT TCAAGACGAT
2251    TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG
        ACACCGCGCC ATAATAGGGC ATAACTGCGG CCCGTTCTCG TTGAGCCAGC
2301    CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG
        GGCGTATGTG ATAAGAGTCT TACTGAACCA ACTCATGAGT GGTCAGTGTC
2351    AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC
        TTTTCGTAGA ATGCCTACCG TACTGTCATT CTCTTAATAC GTCACGACGG
2401    ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA CAACGATCGG
        TATTGGTACT CACTATTGTG ACGCCGGTTG AATGAAGACT GTTGCTAGCC
2451    AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
        TCCTGGCTTC CTCGATTGGC GAAAAAACGT GTTGTACCCC CTAGTACATT
2501    CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC
        GAGCGGAACT AGCAACCCTT GGCCTCGACT TACTTCGGTA TGGTTTGCTG
2551    GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT
        CTCGCACTGT GGTGCTACGG ACATCGTTAC CGTTGTTGCA ACGCGTTTGA
2601    ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT
        TAATTGACCG CTTGATGAAT GAGATCGAAG GGCCGTTGTT AATTATCTGA
2651    GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG
        CCTACCTCCG CCTATTTCAA CGTCCTGGTG AAGACGCGAG CCGGGAAGGC
2701    GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG
        CGACCGACCA AATAACGACT ATTTAGACCT CGGCCACTCG CACCCAGAGC
2751    CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
        GCCATAGTAA CGTCGTGACC CCGGTCTACC ATTCGGGAGG GCATAGCATC
2801    TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG
        AATAGATGTG CTGCCCCTCA GTCCGTTGAT ACCTACTTGC TTTATCTGTC
2851    ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA
        TAGCGACTCT ATCCACGGAG TGACTAATTC GTAACCATTG ACAGTCTGGT
2901    AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA
        TCAAATGAGT ATATATGAAA TCTAACTAAA TTTTGAAGTA AAAATTAAAT
2951    AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT
        TTTCCTAGAT CCACTTCTAG GAAAAACTAT TAGAGTACTG GTTTTAGGGA
3001    TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA
        ATTGCACTCA AAAGCAAGGT GACTCGCAGT CTGGGGCATC TTTTCTAGTT
3051    AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
        TCCTAGAAGA ACTCTAGGAA AAAAAGACGC GCATTAGACG ACGAACGTTT
```

```
3101  CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA
      GTTTTTTTGG TGGCGATGGT CGCCACCAAA CAAACGGCCT AGTTCTCGAT
3151  CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA
      GGTTGAGAAA AAGGCTTCCA TTGACCGAAG TCGTCTCGCG TCTATGGTTT
3201  TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG
      ATGACAGGAA GATCACATCG GCATCAATCC GGTGGTGAAG TTCTTGAGAC
3251  TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
      ATCGTGGCGG ATGTATGGAG CGAGACGATT AGGACAATGG TCACCGACGA
3301  GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT
      CGGTCACCGC TATTCAGCAC AGAATGGCCC AACCTGAGTT CTGCTATCAA
3351  ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
      TGGCCTATTC CGCGTCGCCA GCCCGACTTG CCCCCCAAGC ACGTGTGTCG
3401  CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG
      GGTCGAACCT CGCTTGCTGG ATGTGGCTTG ACTCTATGGA TGTCGCACTC
3451  CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC
      GATACTCTTT CGCGGTGCGA AGGGCTTCCC TCTTTCCGCC TGTCCATAGG
3501  GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG
      CCATTCGCCG TCCCAGCCTT GTCCTCTCGC GTGCTCCCTC GAAGGTCCCC
3551  GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT
      CTTTGCGGAC CATAGAAATA TCAGGACAGC CCAAAGCGGT GGAGACTGAA
3601  GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA
      CTCGCAGCTA AAAACACTAC GAGCAGTCCC CCCGCCTCGG ATACCTTTTT
3651  CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
      GCGGTCGTTG CGCCGGAAAA ATGCCAAGGA CCGGAAAACG ACCGGAAAAC
3701  CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT
      GAGTGTACAA GAAAGGACGC AATAGGGGAC TAAGACACCT ATTGGCATAA
3751  ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG
      TGGCGGAAAC TCACTCGACT ATGGCGAGCG GCGTCGGCTT GCTGGCTCGC
3801  CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA CGCAAACCGC
      GTCGCTCAGT CACTCGCTCC TTCGCCTTCT CGCGGGTTAT GCGTTTGGCG
3851  CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
      GAGAGGGGCG CGCAACCGGC TAAGTAATTA CGTCGACCGT GCTGTCCAAA
3901  CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT
      GGGCTGACCT TTCGCCCGTC ACTCGCGTTG CGTTAATTAC ACTCAATCGA
3951  CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
      GTGAGTAATC CGTGGGGTCC GAAATGTGAA ATACGAAGGC CGAGCATACA
4001  TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC
      ACACACCTTA ACACTCGCCT ATTGTTAAAG TGTGTCCTTT GTCGATACTG
4051  CATGATTACG AATTGAATTG CGGCCGCAAT TCTGAATGTT AAATGTTATA
      GTACTAATGC TTAACTTAAC GCCGGCGTTA AGACTTACAA TTTACAATAT
4101  CTTTGGATGA AGCTATAAAT ATGCATTGGA AAAATAATCC ATTTAAAGAA
      GAAACCTACT TCGATATTTA TACGTAACCT TTTTATTAGG TAAATTTCTT
4151  AGGATTCAAA TACTACAAAA CCTAAGCGAT AATATGTTAA CTAAGCTTAT
      TCCTAAGTTT ATGATGTTTT GGATTCGCTA TTATACAATT GATTCGAATA
4201  TCTTAACGAC GCTTTAAATA TACACAAATA AACATAATTT TTGTATAACC
      AGAATTGCTG CGAAATTTAT ATGTGTTTAT TTGTATTAAA AACATATTGG
4251  TAACAAATAA CTAAAACATA AAAATAATAA AAGGAAATGT AATATCGTAA
      ATTGTTTATT GATTTTGTAT TTTTATTATT TTCCTTTACA TTATAGCATT
4301  TTATTTTACT CAGGAATGGG GTTAAATATT TATATCACGT GTATATCTAT
      AATAAAATGA GTCCTTACCC CAATTTATAA ATATAGTGCA CATATAGATA
4351  ACTGTTATCG TATACTCTTT ACAATTACTA TTACGAATAT GCAAGAGATA
      TGACAATAGC ATATGAGAAA TGTTAATGAT AATGCTTATA CGTTCTCTAT
4401  ATAAGATTAC GTATTTAAGA GAATCTTGTC ATGATAATTG GGTACGACAT
      TATTCTAATG CATAAATTCT CTTAGAACAG TACTATTAAC CCATGCTGTA
4451  AGTGATAAAT GCTATTTCGC ATCGTTACAT AAAGTCAGTT GGAAAGATGG
      TCACTATTTA CGATAAAGCG TAGCAATGTA TTTCAGTCAA CCTTTCTACC
4501  ATTTGACAGA TGTAACTTAA TAGGTGCAAA AATGTTAAAT AACAGCATTC
      TAAACTGTCT ACATTGAATT ATCCACGTTT TTACAATTTA TTGTCGTAAG
4551  TATCGGAAGA TAGGATACCA GTTATATTAT ACAAAAATCA CTGGTTGGAT
```

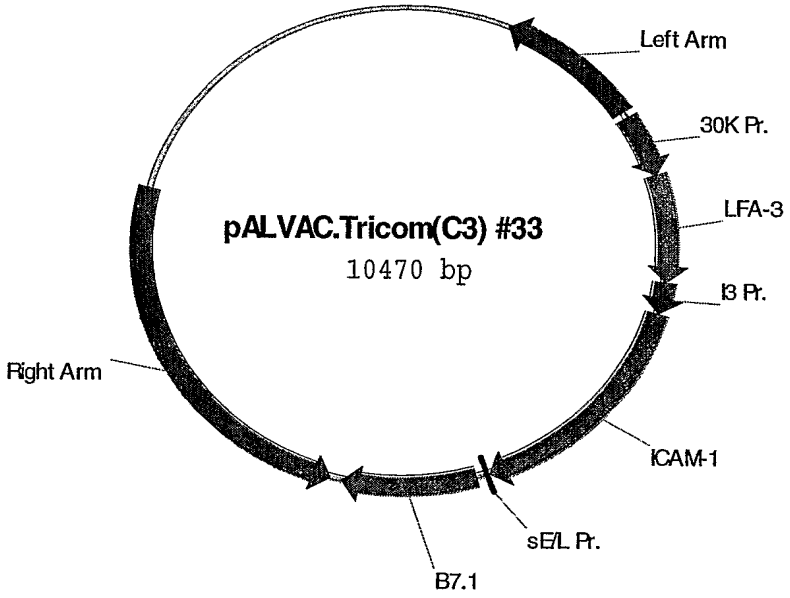
```
       ATAGCCTTCT ATCCTATGGT CAATATAATA TGTTTTTAGT GACCAACCTA
4601   AAAACAGATT CTGCAATATT CGTAAAAGAT GAAGATTACT GCGAATTTGT
       TTTTGTCTAA GACGTTATAA GCATTTTCTA CTTCTAATGA CGCTTAAACA
4651   AAACTATGAC AATAAAAAGC CATTTATCTC AACGACATCG TGTAATTCTT
       TTTGATACTG TTATTTTTCG GTAAATAGAG TTGCTGTAGC ACATTAAGAA
4701   CCATGTTTTA TGTATGTGTT TCAGATATTA TGAGATTACT ATAAACTTTT
       GGTACAAAAT ACATACACAA AGTCTATAAT ACTCTAATGA TATTTGAAAA
4751   TGTATACTTA TATTCCGTAA ACTATATTAA TCATGAAGAA AATGAAAAAG
       ACATATGAAT ATAAGGCATT TGATATAATT AGTACTTCTT TTACTTTTTC
4801   TATAGAAGCT GTTCACGAGC GGTTGTTGAA AACAACAAAA TTATACATTC
       ATATCTTCGA CAAGTGCTCG CCAACAACTT TTGTTGTTTT AATATGTAAG
4851   AAGATGGCTT ACATATACGT CTGTGAGGCT ATCATGGATA ATGACAATGC
       TTCTACCGAA TGTATATGCA GACACTCCGA TAGTACCTAT TACTGTTACG
4901   ATCTCTAAAT AGGTTTTTGG ACAATGGATT CGACCCTAAC ACGGAATATG
       TAGAGATTTA TCCAAAAACC TGTTACCTAA GCTGGGATTG TGCCTTATAC
4951   GTACTCTACA ATCTCCTCTT GAAATGGCTG TAATGTTCAA GAATACCGAG
       CATGAGATGT TAGAGGAGAA CTTTACCGAC ATTACAAGTT CTTATGGCTC
5001   GCTATAAAAA TCTTGATGAG GTATGGAGCT AAACCTGTAG TTACTGAATG
       CGATATTTTT AGAACTACTC CATACCTCGA TTTGGACATC AATGACTTAC
5051   CACAACTTCT TGTCTGCATG ATGCGGTGTT GAGAGACGAC TACAAAATAG
       GTGTTGAAGA ACAGACGTAC TACGCCACAA CTCTCTGCTG ATGTTTTATC
5101   TGAAAGATCT GTTGAAGAAT AACTATGTAA ACAATGTTCT TTACAGCGGA
       ACTTTCTAGA CAACTTCTTA TTGATACATT TGTTACAAGA AATGTCGCCT
5151   GGCTTTACTC CTTTGTGTTT GGCAGCTTAC CTTAACAAAG TTAATTTGGT
       CCGAAATGAG GAAACACAAA CCGTCGAATG GAATTGTTTC AATTAAACCA
5201   TAAACTTCTA TTGGCTCATT CGGCGGATGT AGATATTTCA AACACGGATC
       ATTTGAAGAT AACCGAGTAA GCCGCCTACA TCTATAAAGT TTGTGCCTAG
5251   GGTTAACTCC TCTACATATA GCCGTATCAA ATAAAAATTT AACAATGGTT
       CCAATTGAGG AGATGTATAT CGGCATAGTT TATTTTTAAA TTGTTACCAA
5301   AAACTTCTAT TGAACAAAGG TGCTGATACT GACTTGCTGG ATAACATGGG
       TTTGAAGATA ACTTGTTTCC ACGACTATGA CTGAACGACC TATTGTACCC
5351   ATGTACTCCT TTAATGATCG CTGTACAATC TGGAAATATT GAAATATGTA
       TACATGAGGA AATTACTAGC GACATGTTAG ACCTTTATAA CTTTATACAT
5401   GCACACTACT TAAAAAAAAT AAAATGTCCA GAACTGGGAA AAATTGATCT
       CGTGTGATGA ATTTTTTTTA TTTTACAGGT CTTGACCCTT TTTAACTAGA
5451   TGCCAGCTGT AATTCATGGT AGAAAAGAAG TGCTCAGGCT ACTTTTCAAC
       ACGGTCGACA TTAAGTACCA TCTTTTCTTC ACGAGTCCGA TGAAAAGTTG
5501   AAAGGAGCAG ATGTAAACTA CATCTTTGAA AGAAATGGAA AATCATATAC
       TTTCCTCGTC TACATTTGAT GTAGAAACTT TCTTTACCTT TTAGTATATG
5551   TGTTTTGGAA TTGATTAAAG AAAGTTACTC TGAGACACAA AAGAGGTAGC
       ACAAAACCTT AACTAATTTC TTTCAATGAG ACTCTGTGTT TTCTCCATCG
5601   TGAAGTGGTA CTCTCAAAGG TACGTGACTA ATTAGCTATA AAAAGGATCC
       ACTTCACCAT GAGAGTTTCC ATGCACTGAT TAATCGATAT TTTTCCTAGG
5651   TAGAGGATCA TTATTTAACG TAAACTAAAT GGAAAAGCTA TTTACAGGTA
       ATCTCCTAGT AATAAATTGC ATTTGATTTA CCTTTTCGAT AAATGTCCAT
5701   CATACGGTGT TTTCTGGAAT CAAATGATTC TGATTTTGAG GATTTTATCA
       GTATGCCACA AAAGACCTTA GTTTACTAAG ACTAAAACTC CTAAAATAGT
5751   ATACAATAAT GACAGTGCTA ACTGGTAAAA AAGAAAGCAA ACAATTATCA
       TATGTTATTA CTGTCACGAT TGACCATTTT TTCTTTCGTT TGTTAATAGT
5801   TGGCTAACAA TTTTTATTAT ATTTGTAGTA TGCATAGTGG TCTTTACGTT
       ACCGATTGTT AAAAATAATA TAAACATCAT ACGTATCACC AGAAATGCAA
5851   TCTTTATTTA AAGTTAATGT GTTAAGATTA AATGGAGTAA TTGGATCCCC
       AGAAATAAAT TTCAATTACA CAATTCTAAT TTACCTCATT AACCTAGGGG
5901   CATCGATGGG GAATTCACTG GCCGTCGTTT TACAACGTCG TGACTGGGAA
       GTAGCTACCC CTTAAGTGAC CGGCAGCAAA ATGTTGCAGC ACTGACCCTT
5951   AACCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTCGC
       TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG
6001   CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT
       GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA AGGGTTGTCA
```

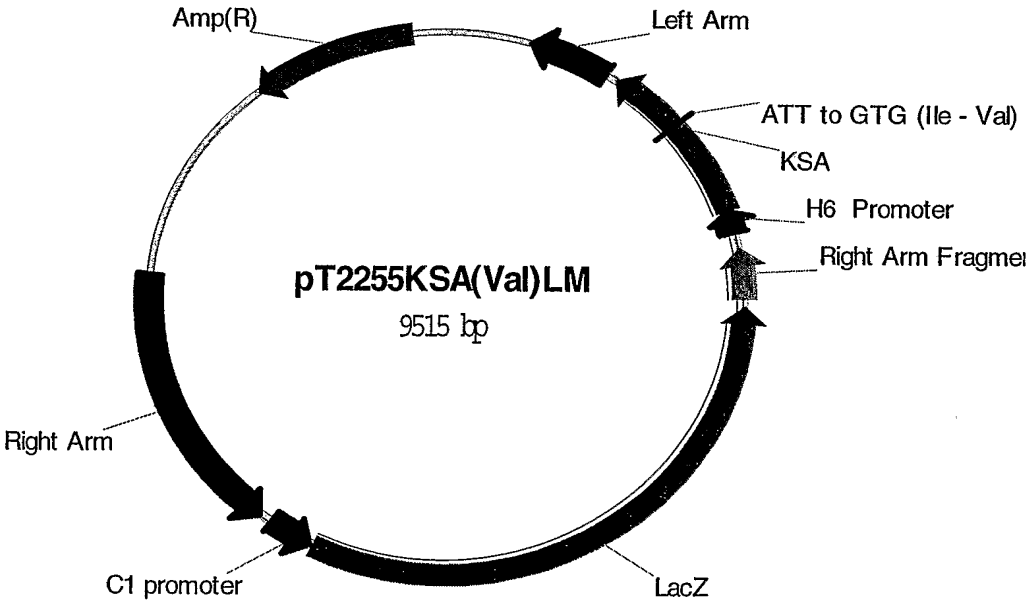
```
6051  TGCGCAGCCT GAATGGCGAA TGGCGCTTTG CCTGGTTTCC GGCACCAGAA
      ACGCGTCGGA CTTACCGCTT ACCGCGAAAC GGACCAAAGG CCGTGGTCTT
6101  GCGGTGCCGG AAAGCTGGCT GGAGTGCGAT CTTCCTGAGG CCGATACTGT
      CGCCACGGCC TTTCGACCGA CCTCACGCTA GAAGGACTCC GGCTATGACA
6151  CGTCGTCCCC TCAAACTGGC AGATGCACGG TTACGATGCG CCCATCTACA
      GCAGCAGGGG AGTTTGACCG TCTACGTGCC AATGCTACGC GGGTAGATGT
6201  CCAACGTAAC CTATCCCATT ACGGTCAATC CGCCGTTTGT TCCCACGGAG
      GGTTGCATTG GATAGGGTAA TGCCAGTTAG GCGGCAAACA AGGGTGCCTC
6251  AATCCGACGG GTTGTTACTC GCTCACATTT AATGTTGATG AAAGCTGGCT
      TTAGGCTGCC CAACAATGAG CGAGTGTAAA TTACAACTAC TTTCGACCGA
6301  ACAGGAAGGC CAGACGCGAA TTATTTTTGA TGGCGTTAAC TCGGCGTTTC
      TGTCCTTCCG GTCTGCGCTT AATAAAAACT ACCGCAATTG AGCCGCAAAG
6351  ATCTGTGGTG CAACGGGCGC TGGGTCGGTT ACGGCCAGGA CAGTCGTTTG
      TAGACACCAC GTTGCCCGCG ACCCAGCCAA TGCCGGTCCT GTCAGCAAAC
6401  CCGTCTGAAT TTGACCTGAG CGCATTTTTA CGCGCCGGAG AAAACCGCCT
      GGCAGACTTA AACTGGACTC GCGTAAAAAT GCGCGGCCTC TTTTGGCGGA
6451  CGCGGTGATG GTGCTGCGTT GGAGTGACGG CAGTTATCTG GAAGATCAGG
      GCGCCACTAC CACGACGCAA CCTCACTGCC GTCAATAGAC CTTCTAGTCC
6501  ATATGTGGCG GATGAGCGGC ATTTTCCGTG ACGTCTCGTT GCTGCATAAA
      TATACACCGC CTACTCGCCG TAAAAGGCAC TGCAGAGCAA CGACGTATTT
6551  CCGACTACAC AAATCAGCGA TTTCCATGTT GCCACTCGCT TTAATGATGA
      GGCTGATGTG TTTAGTCGCT AAAGGTACAA CGGTGAGCGA AATTACTACT
6601  TTTCAGCCGC GCTGTACTGG AGGCTGAAGT TCAGATGTGC GGCGAGTTGC
      AAAGTCGGCG CGACATGACC TCCGACTTCA AGTCTACACG CCGCTCAACG
6651  GTGACTACCT ACGGGTAACA GTTTCTTTAT GGCAGGGTGA AACGCAGGTC
      CACTGATGGA TGCCCATTGT CAAAGAAATA CCGTCCCACT TTGCGTCCAG
6701  GCCAGCGGCA CCGCGCCTTT CGGCGGTGAA ATTATCGATG AGCGTGGTGG
      CGGTCGCCGT GGCGCGGAAA GCCGCCACTT TAATAGCTAC TCGCACCACC
6751  TTATGCCGAT CGCGTCACAC TACGTCTGAA CGTCGAAAAC CCGAAACTGT
      AATACGGCTA GCGCAGTGTG ATGCAGACTT GCAGCTTTTG GGCTTTGACA
6801  GGAGCGCCGA AATCCCGAAT CTCTATCGTG CGGTGGTTGA ACTGCACACC
      CCTCGCGGCT TTAGGGCTTA GAGATAGCAC GCCACCAACT TGACGTGTGG
6851  GCCGACGGCA CGCTGATTGA AGCAGAAGCC TGCGATGTCG GTTTCCGCGA
      CGGCTGCCGT GCGACTAACT TCGTCTTCGG ACGCTACAGC CAAAGGCGCT
6901  GGTGCGGATT GAAAATGGTC TGCTGCTGCT GAACGGCAAG CCGTTGCTGA
      CCACGCCTAA CTTTTACCAG ACGACGACGA CTTGCCGTTC GGCAACGACT
6951  TTCGAGGCGT TAACCGTCAC GAGCATCATC CTCTGCATGG TCAGGTCATG
      AAGCTCCGCA ATTGGCAGTG CTCGTAGTAG GAGACGTACC AGTCCAGTAC
7001  GATGAGCAGA CGATGGTGCA GGATATCCTG CTGATGAAGC AGAACAACTT
      CTACTCGTCT GCTACCACGT CCTATAGGAC GACTACTTCG TCTTGTTGAA
7051  TAACGCCGTG CGCTGTTCGC ATTATCCGAA CCATCCGCTG TGGTACACGC
      ATTGCGGCAC GCGACAAGCG TAATAGGCTT GGTAGGCGAC ACCATGTGCG
7101  TGTGCGACCG CTACGGCCTG TATGTGGTGG ATGAAGCCAA TATTGAAACC
      ACACGCTGGC GATGCCGGAC ATACACCACC TACTTCGGTT ATAACTTTGG
7151  CACGGCATGG TGCCAATGAA TCGTCTGACC GATGATCCGC GCTGGCTACC
      GTGCCGTACC ACGGTTACTT AGCAGACTGG CTACTAGGCG CGACCGATGG
7201  GGCGATGAGC GAACGCGTAA CGCGAATGGT GCAGCGCGAT CGTAATCACC
      CCGCTACTCG CTTGCGCATT GCGCTTACCA CGTCGCGCTA GCATTAGTGG
7251  CGAGTGTGAT CATCTGGTCG CTGGGGAATG AATCAGGCCA CGGCGCTAAT
      GCTCACACTA GTAGACCAGC GACCCCTTAC TTAGTCCGGT GCCGCGATTA
7301  CACGACGCGC TGTATCGCTG GATCAAATCT GTCGATCCTT CCCGCCCGGT
      GTGCTGCGCG ACATAGCGAC CTAGTTTAGA CAGCTAGGAA GGGCGGGCCA
7351  GCAGTATGAA GGCGGCGGAG CCGACACCAC GGCCACCGAT ATTATTTGCC
      CGTCATACTT CCGCCGCCTC GGCTGTGGTG CCGGTGGCTA TAATAAACGG
7401  CGATGTACGC GCGCGTGGAT GAAGACCAGC CCTTCCCGGC TGTGCCGAAA
      GCTACATGCG CGCGCACCTA CTTCTGGTCG GGAAGGGCCG ACACGGCTTT
7451  TGGTCCATCA AAAAATGGCT TTCGCTACCT GGAGAGACGC GCCCGCTGAT
      ACCAGGTAGT TTTTTACCGA AAGCGATGGA CCTCTCTGCG CGGGCGACTA
7501  CCTTTGCGAA TACGCCCACG CGATGGGTAA CAGTCTTGGC GGTTTCGCTA
```

```
        GGAAACGCTT ATGCGGGTGC GCTACCCATT GTCAGAACCG CCAAAGCGAT
7551    AATACTGGCA GGCGTTTCGT CAGTATCCCC GTTTACAGGG CGGCTTCGTC
        TTATGACCGT CCGCAAAGCA GTCATAGGGG CAAATGTCCC GCCGAAGCAG
7601    TGGGACTGGG TGGATCAGTC GCTGATTAAA TATGATGAAA ACGGCAACCC
        ACCCTGACCC ACCTAGTCAG CGACTAATTT ATACTACTTT TGCCGTTGGG
7651    GTGGTCGGCT TACGGCGGTG ATTTTGGCGA TACGCCGAAC GATCGCCAGT
        CACCAGCCGA ATGCCGCCAC TAAAACCGCT ATGCGGCTTG CTAGCGGTCA
7701    TCTGTATGAA CGGTCTGGTC TTTGCCGACC GCACGCCGCA TCCAGCGCTG
        AGACATACTT GCCAGACCAG AAACGGCTGG CGTGCGGCGT AGGTCGCGAC
7751    ACGGAAGCAA AACACCAGCA GCAGTTTTTC CAGTTCCGTT TATCCGGGCA
        TGCCTTCGTT TTGTGGTCGT CGTCAAAAAG GTCAAGGCAA ATAGGCCCGT
7801    AACCATCGAA GTGACCAGCG AATACCTGTT CCGTCATAGC GATAACGAGC
        TTGGTAGCTT CACTGGTCGC TTATGGACAA GGCAGTATCG CTATTGCTCG
7851    TCCTGCACTG GATGGTGGCG CTGGATGGTA AGCCGCTGGC AAGCGGTGAA
        AGGACGTGAC CTACCACCGC GACCTACCAT TCGGCGACCG TTCGCCACTT
7901    GTGCCTCTGG ATGTCGCTCC ACAAGGTAAA CAGTTGATTG AACTGCCTGA
        CACGGAGACC TACAGCGAGG TGTTCCATTT GTCAACTAAC TTGACGGACT
7951    ACTACCGCAG CCGGAGAGCG CCGGGCAACT CTGGCTCACA GTACGCGTAG
        TGATGGCGTC GGCCTCTCGC GGCCCGTTGA GACCGAGTGT CATGCGCATC
8001    TGCAACCGAA CGCGACCGCA TGGTCAGAAG CCGGGCACAT CAGCGCCTGG
        ACGTTGGCTT GCGCTGGCGT ACCAGTCTTC GGCCCGTGTA GTCGCGGACC
8051    CAGCAGTGGC GTCTGGCGGA AAACCTCAGT GTGACGCTCC CCGCCGCGTC
        GTCGTCACCG CAGACCGCCT TTTGGAGTCA CACTGCGAGG GGCGGCGCAG
8101    CCACGCCATC CCGCATCTGA CCACCAGCGA AATGGATTTT TGCATCGAGC
        GGTGCGGTAG GGCGTAGACT GGTGGTCGCT TTACCTAAAA ACGTAGCTCG
8151    TGGGTAATAA GCGTTGGCAA TTTAACCGCC AGTCAGGCTT TCTTTCACAG
        ACCCATTATT CGCAACCGTT AAATTGGCGG TCAGTCCGAA AGAAAGTGTC
8201    ATGTGGATTG GCGATAAAAA ACAACTGCTG ACGCCGCTGC GCGATCAGTT
        TACACCTAAC CGCTATTTTT TGTTGACGAC TGCGGCGACG CGCTAGTCAA
8251    CACCCGTGCA CCGCTGGATA ACGACATTGG CGTAAGTGAA GCGACCCGCA
        GTGGGCACGT GGCGACCTAT TGCTGTAACC GCATTCACTT CGCTGGGCGT
8301    TTGACCCTAA CGCCTGGGTC GAACGCTGGA AGGCGGCGGG CCATTACCAG
        AACTGGGATT GCGGACCCAG CTTGCGACCT TCCGCCGCCC GGTAATGGTC
8351    GCCGAAGCAG CGTTGTTGCA GTGCACGGCA GATACACTTG CTGATGCGGT
        CGGCTTCGTC GCAACAACGT CACGTGCCGT CTATGTGAAC GACTACGCCA
8401    GCTGATTACG ACCGCTCACG CGTGGCAGCA TCAGGGGAAA ACCTTATTTA
        CGACTAATGC TGGCGAGTGC GCACCGTCGT AGTCCCCTTT TGGAATAAAT
8451    TCAGCCGGAA AACCTACCGG ATTGATGGTA GTGGTCAAAT GGCGATTACC
        AGTCGGCCTT TTGGATGGCC TAACTACCAT CACCAGTTTA CCGCTAATGG
8501    GTTGATGTTG AAGTGGCGAG CGATACACCG CATCCGGCGC GGATTGGCCT
        CAACTACAAC TTCACCGCTC GCTATGTGGC GTAGGCCGCG CCTAACCGGA
8551    GAACTGCCAG CTGGCGCAGG TAGCAGAGCG GGTAAACTGG CTCGGATTAG
        CTTGACGGTC GACCGCGTCC ATCGTCTCGC CCATTTGACC GAGCCTAATC
8601    GGCCGCAAGA AAACTATCCC GACCGCCTTA CTGCCGCCTG TTTTGACCGC
        CCGGCGTTCT TTTGATAGGG CTGGCGGAAT GACGGCGGAC AAAACTGGCG
8651    TGGGATCTGC CATTGTCAGA CATGTATACC CCGTACGTCT TCCCGAGCGA
        ACCCTAGACG GTAACAGTCT GTACATATGG GGCATGCAGA AGGGCTCGCT
8701    AAACGGTCTG CGCTGCGGGA CGCGCGAATT GAATTATGGC CCACACCAGT
        TTTGCCAGAC GCGACGCCCT GCGCGCTTAA CTTAATACCG GGTGTGGTCA
8751    GGCGCGGCGA CTTCCAGTTC AACATCAGCC GCTACAGTCA ACAGCAACTG
        CCGCGCCGCT GAAGGTCAAG TTGTAGTCGG CGATGTCAGT TGTCGTTGAC
8801    ATGGAAACCA GCCATCGCCA TCTGCTGCAC GCGGAAGAAG GCACATGGCT
        TACCTTTGGT CGGTAGCGGT AGACGACGTG CGCCTTCTTC CGTGTACCGA
8851    GAATATCGAC GGTTTCCATA TGGGGATTGG TGGCGACGAC TCCTGGAGCC
        CTTATAGCTG CCAAAGGTAT ACCCCTAACC ACCGCTGCTG AGGACCTCGG
8901    CGTCAGTATC GGCGGAATTC CAGCTGAGCG CCGGTCGCTA CCATTACCAG
        GCAGTCATAG CCGCCTTAAG GTCGACTCGC GGCCAGCGAT GGTAATGGTC
8951    TTGGTCTGGT GTCAAAAATA ATAATAACCG GGCAGGGGGG ATCCGGAGCT
        AACCAGACCA CAGTTTTTAT TATTATTGGC CCGTCCCCCC TAGGCCTCGA
```

```
9001  TATCGCAGAT CAATGATCGC TGTACAATCT GGAAATATTG AAATATGTAG
      ATAGCGTCTA GTTACTAGCG ACATGTTAGA CCTTTATAAC TTTATACATC
9051  CACACTACTT AAAAAAAATA AAATGTCCAG AACTGGGAAA AATTGATCTT
      GTGTGATGAA TTTTTTTTAT TTTACAGGTC TTGACCCTTT TTAACTAGAA
9101  GCCAGCTGTA ATTCATGGTA GAAAAGAAGT GCTCAGGCTA CTTTTCAACA
      CGGTCGACAT TAAGTACCAT CTTTTCTTCA CGAGTCCGAT GAAAAGTTGT
9151  AAGGAGCAGA TGTAAACTAC ATCTTTGAAA GAAATGGAAA ATCATATACT
      TTCCTCGTCT ACATTTGATG TAGAAACTTT CTTTACCTTT TAGTATATGA
9201  GTTTTGGAAT TGATTAAAGA AAGTTACTCT GAGACACAAA AGAGGTAGCT
      CAAAACCTTA ACTAATTTCT TTCAATGAGA CTCTGTGTTT TCTCCATCGA
9251  GAAGTGGTAC TCTCAAAGGT ACGTGACTAA TTAGCTATAA AAAGGATCCG
      CTTCACCATG AGAGTTTCCA TGCACTGATT AATCGATATT TTTCCTAGGC
9301  GTACCCTCGA GTCTAGAATC GATCCCGGGT TAATTAATTA GTTATTAGAC
      CATGGGAGCT CAGATCTTAG CTAGGGCCCA ATTAATTAAT CAATAATCTG
9351  AAGGTGAAAA CGAAACTATT TGTAGCTTAA TTAATTAGAG CTTCTTTATT
      TTCCACTTTT GCTTTGATAA ACATCGAATT AATTAATCTC GAAGAAATAA
9401  CTATACTTAA AAAGTGAAAA TAAATACAAA GGTTCTTGAG GGTTGTGTTA
      GATATGAATT TTTCACTTTT ATTTATGTTT CCAAGAACTC CCAACACAAT
9451  AATTGAAAGC GAGAAATAAT CATAAATTAT TTCATTATCG CGATATCCGT
      TTAACTTTCG CTCTTTATTA GTATTTAATA AAGTAATAGC GCTATAGGCA
9501  TAAGTTTGTA TCGTA
      ATTCAAACAT AGCAT
```

FIGURE 6


Left Arm
30K Pr.
LFA-3
I3 Pr.
ICAM-1
sE/L Pr.
B7.1
Right Arm
pALVAC.Tricom(C3) #33
10470 bp


Amp(R)
Left Arm
ATT to GTG (Ile - Val)
KSA
H6 Promoter
Right Arm Fragme
pT2255KSA(Val)LM
9515 bp
Right Arm
C1 promoter
LacZ